ISBN: 978-84-7666-194-9

Avances y desafíos en la evaluación educativa

En las últimas décadas se han producido notables avances en el ámbito de la evaluación. Sin embargo, son todavía escasos los sistemas educativos que cuentan con modelos asentados de evaluación de escuelas y de la práctica docente. La evaluación de los aprendizajes en el aula necesita cambios profundos. La evaluación de las propias políticas educativas y de los programas de innovación que desde ellas se impulsan tiene también poca tradición en nuestro entorno. Este libro ofrece una panorámica de los diversos niveles y ámbitos de la evaluación, destacando los avances y los retos aún pendientes.

Series de la colección

Alfabetización	Cultura escrita	Educación artística
Educación técnico-profesional	**Evaluación**	Infancia
Profesión docente	Reformas educativas	TIC

Evaluación

La evaluación de los sistemas educativos, de los programas, de las escuelas, de los docentes y de los alumnos es una condición necesaria para mejorar la calidad de la educación. La serie de libros sobre **evaluación** pretende ofrecer los principales modelos, indicadores, experiencias e innovaciones que se desarrollan en este campo.

Metas Educativas 2021

La conmemoración de los bicentenarios de las independencias debe favorecer una iniciativa capaz de generar un gran apoyo colectivo. Así lo entendieron los ministros de Educación iberoamericanos cuando respaldaron de forma unánime el proyecto **Metas Educativas 2021: la educación que queremos para la generación de los Bicentenarios.** Semejante tarea colectiva, articulada en torno a la educación, ha de contribuir al desarrollo económico y social de la región y a la formación de ciudadanos cultos y libres en sociedades justas y democráticas. La ***Colección Metas Educativas 2021*** pretende ampliar y compartir el conocimiento e impulsar el debate, la participación y el compromiso colectivo con este ambicioso proyecto.

MINISTERIO DE ASUNTOS EXTERIORES Y DE COOPERACIÓN
aecid

FundaciónSantillana
Avances y desafíos en la evaluación educativa
Evaluación
Metas Educativas 2021
OEI

Evaluación

Avances y desafíos en la evaluación educativa

Elena Martín
Felipe Martínez Rizo
Coordinadores

Metas Educativas 2021

La educación que queremos para la generación de los Bicentenarios

Organização dos Estados Ibero-americanos
Para a Educação, a Ciência e a Cultura
OEI
Organización de Estados Iberoamericanos
Para la Educación, la Ciencia y la Cultura

FundaciónSantillana

134027_C.indd 1
Negro Pantone 375 C
13/3/09 21:17:30

Evaluación

Avances y desafíos en la evaluación educativa

Elena Martín
Felipe Martínez Rizo
Coordinadores

Organização dos Estados Ibero-americanos
Para a Educação, a Ciência e a Cultura

OEI

Organización de Estados Iberoamericanos
Para la Educación, la Ciencia y la Cultura

FundaciónSantillana

© Del texto: Organización de Estados Iberoamericanos
para la Educación, la Ciencia y la Cultura (OEI)
C/ Bravo Murillo, 38
28015 Madrid, España
www.oei.es

Las opiniones de los autores expresadas en este libro no representan necesariamente los puntos de vista de la OEI.

La colección METAS EDUCATIVAS 2021 es una iniciativa de la OEI en colaboración con la Fundación Santillana.

Impreso en España por

ISBN: 978-84-7666-194-9
Depósito legal:

Índice

Preámbulo, *Álvaro Marchesi* 7

Introducción, *Elena Martín y Felipe Martínez Rizo* 11

La evaluación educativa hoy 15

- Evaluación y cambio educativo: los debates actuales sobre las ventajas y los riesgos de la evaluación, *Alejandro Tiana* 17
- La evaluación de la calidad de los sistemas educativos: propuesta de un modelo, *Felipe Martínez Rizo* 27
- Las evaluaciones internacionales, *Enrique Roca* 41
- Las evaluaciones regionales y nacionales en América Latina, *Guillermo Ferrer* 53

Las múltiples caras de la evaluación 65

- Indicadores y desafíos de los sistemas educativos en América Latina, *Margarita Poggi* 67
- La evaluación de políticas y programas: desafíos para la gestión en educación, *Marta Kisilevsky* 81
- Currículo y evaluación estandarizada: colaboración o tensión, *Elena Martín* 89
- La evaluación de centros: riesgos y posibilidades, *Leda Badilla* 99
- La evaluación del desempeño docente para el desarrollo de las competencias profesionales, *Pedro Ravela* 113
- Evaluación de los aprendizajes en el aula: una conceptualización renovada, *Juan Manuel Esquivel* 127

El sentido de la evaluación 145

- Difusión y uso de resultados de evaluaciones educativas a gran escala en América Latina, *Patricia McLauchlan de Arregui* 147
- Los resultados de las evaluaciones y su papel en las políticas educativas, *Álvaro Marchesi* 161

Bibliografía 169

Los autores 179

Preámbulo

Álvaro Marchesi
Secretario general de la OEI

En las dos últimas décadas se ha producido un enorme avance de la reflexión teórica y de los programas aplicados en el campo de la evaluación, sobre todo, en el ámbito de la evaluación externa y estandarizada. Las aportaciones no han sido tan claras, sin embargo, en lo referente a la evaluación de centros, ni ha llegado con igual fuerza a la evaluación en el aula. No obstante, existe un amplio consenso en todos los países de la región en que la evaluación sistemática y correctamente orientada constituye una de las estrategias principales para la mejora de la calidad de la enseñanza. Fruto de este convencimiento es la creación en la mayoría de los países de Institutos Nacionales responsables de la evaluación educativa, el desarrollo de diversos sistemas de indicadores, la creciente participación de los países en estudios internacionales como PISA, dirigido por la OCDE, o aquellos coordinados por la IEA, así como los estudios llevados a cabo por OREALC/UNESCO a través del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE), cuyas tareas se realizan junto con los organismos responsables de la evaluación de los Ministerios de Educación de cada país.

Tales progresos han servido para destacar que la evaluación debe enfrentarse a dilemas importantes que no son sencillos de resolver, que existen determinadas líneas de acción que deben ser tenidas en cuenta, y que cualquier avance abre nuevos desafíos que exigen reforzar la reflexión teórica, el análisis de las prácticas, la coordinación institucional y la cooperación internacional.

El conflicto existente entre los objetivos de la educación y los objetivos de la evaluación afecta posiblemente a la mayoría de los debates que en la actualidad se manifiestan en este campo. A veces da la impresión de que las principales metas de la educación –conocer, convivir, hacer, ser– y determinadas competencias y valores asociados a ellas –innovación, creatividad, compasión, sensibilidad, justicia– apenas tienen que ver con los proyectos de evaluación y con la información obtenida a través de ellos o, al menos, con aquella información que se transmite a la opinión pública. Aunque los estudios PISA establecen no solo el lugar que cada país alcanza en los logros académicos de sus alumnos, sino también la equidad de su sistema escolar, prácticamente todos los medios de comunicación han destacado en sus titulares la primera de las informaciones y han relegado a un segundo plano la segunda.

Esta tensión entre el sentido de la educación y el sentido de la evaluación está presente en la mayoría de los ámbitos educativos. Se muestra en la evaluación del sistema educativo y de sus programas, así como en la evaluación de las escuelas, de los docentes y de los alumnos. Es cierto que el rigor de los sistemas de obtención de información y de análisis, exigidos por la mayoría de las evaluaciones, contrasta con la libertad del discurso pedagógico, lo que puede explicar buena parte de los desajustes existentes. Pero no es menos cierto que, en ocasiones, da la impresión de que cada uno de los discursos sigue trayectorias claramente diferentes, expresión en muchos casos de opciones ideológicas distintas. A veces, las evaluaciones no pueden, por razones metodológicas, ajustarse a las intenciones educativas establecidas. Otras, por el contrario, pretenden orientarlas y

modificarlas, y establecer, a través de la evaluación, lo que deben ser los objetivos prioritarios de la educación escolar.

Hay que tener en cuenta la diversidad de situaciones educativas en cada país, así como las diferencias existentes en el seno de cada uno de ellos. Este panorama complica no solo las evaluaciones internacionales, sino también los propios sistemas nacionales de evaluación. Los procesos de descentralización y de apuesta por la autonomía de las escuelas obligan a adaptar los sistemas de evaluación a la nueva realidad educativa. Lo mismo sucede, o debería suceder, si se tiene en cuenta la diversidad de culturas y de lenguas presentes en Iberoamérica, no digamos la diversidad del alumnado. El desafío para los sistemas y proyectos de evaluación es lograr un acertado equilibrio entre los modelos comunes de evaluación y su correcta adaptación a la enorme pluralidad existente, y evitar de esa forma que el interés por mejorar la gestión educativa a través de la descentralización y de la autonomía no termine por reforzar la homogeneización a través de la evaluación.

No hay que olvidar que las evaluaciones se diseñan desde una perspectiva ideológica y educativa determinada, y que los datos que se obtienen en ellas no permiten obtener conclusiones inmediatas y directas. Es importante destacar que no puede olvidarse en la labor interpretativa la influencia del contexto y de la situación de partida de los sistemas educativos, de los programas y de las escuelas evaluadas; el impacto de la organización y del funcionamiento de las escuelas; el efecto del proceso de enseñanza y de aprendizaje en el aula y de los factores que inciden en el desarrollo del currículo, incluso el propio diseño de currículo y las intenciones educativas contenidas en él; y la influencia, finalmente, de las políticas educativas en marcha y de la propia gestión de las administraciones educativas. Lo que se muestra de forma cada vez más clara es que cualquier evaluación, por limitado que sea su alcance, no debe prescindir de un enfoque sistémico de la realidad educativa, inmersa además en un contexto social determinado.

Todo ello pone de relieve el cuidado con el que hay que abordar las evaluaciones educativas y la utilización de los resultados que se obtienen. La presentación comparada de los resultados directos que logran los alumnos en las pruebas de rendimiento no parece que sea una estrategia sensible al conjunto de factores que condicionan los resultados obtenidos. Tampoco es capaz de reconocer el posible esfuerzo que la escuela y los profesores han podido realizar para ir más allá de lo que se esperaría de los alumnos a partir de su contexto, de sus condiciones iniciales y de sus conocimientos previos.

La evaluación del profesorado se configura en estos momentos como uno de los grandes retos que deben ser abordados con el máximo cuidado. Sin duda, la formación y el desempeño de los docentes es uno de los factores principales de la calidad de un sistema educativo, y la evaluación de su acción educadora puede llegar a ser una positiva estrategia de apoyo a los docentes y de mejora de la calidad de la enseñanza. Pero es preciso tener en cuenta que el trabajo de los profesores no es ajeno a su formación inicial y permanente, a sus condiciones laborales, a la existencia de una carrera profesional y al contexto social y educativo en el que desempeñan su trabajo. Por ello, las propuestas de evaluación del trabajo de los profesores no deben olvidarse del contexto en el que desarrollan su labor y han de estar estrechamente relacionadas con el desarrollo de su carrera profesional y con el apoyo para mejorar la calidad de su enseñanza.

Lo retos a los que se enfrenta la evaluación educativa deben servir de estímulo para la incorporación de metodologías variadas que den cuenta, desde sus diferentes aproximaciones, de la complejidad de la acción educadora, tanto de las instituciones como de sus profesionales. Por

ello, junto con los enfoques más cuantitativos, deberían tenerse en cuenta otros de corte cualitativo que ayudaran a la interpretación de los datos obtenidos. De la misma forma, el énfasis en el diálogo con aquellos que van a participar en los procesos de evaluación, la información sobre las características del mismo y sobre sus consecuencias, así como el cuidado para que se sientan participantes activos y no solo sujetos pacientes de programas diseñados a sus espaldas, son condiciones que garantizan un mayor éxito de las evaluaciones, es decir, un impacto positivo en la mejora de la calidad de la educación.

Hay que reconocer que queda todavía mucho camino por recorrer. Es preciso planificar las políticas de acuerdo con una sólida fundamentación teórica que aproveche lo que se ha podido aprender de todas las experiencias que vienen realizándose en las últimas décadas y ser capaz de enfrentarse a los dilemas existentes en este campo. Además, no hay que perder de vista que los objetivos de la evaluación son mejorar la calidad de la educación, lograr reducir las desigualdades educativas, orientar el cambio educativo y las políticas públicas que lo gestionan, fortalecer las escuelas, favorecer el desarrollo profesional de los docentes y mejorar el aprendizaje de todos los alumnos. Todo ello constituye, sin duda, una agenda enormemente exigente.

En este contexto de progresos y de nuevos desafíos, la OEI pretende colaborar, junto con los países y los demás organismos internacionales presentes en la región, en el impulso a la reflexión teórica, en el apoyo a las iniciativas nacionales e internacionales en el campo de la evaluación, en el fortalecimiento de las instituciones de evaluación presentes en cada país y en la búsqueda de una mayor relación entre las finalidades de la educación y el desarrollo de las futuras evaluaciones. Existe una oportunidad histórica para lograrlo en el desarrollo del proyecto "Metas Educativas 2021: la educación que queremos para la generación de los Bicentenarios". Entre sus objetivos para transformar la educación a lo largo de la próxima década se encuentran los referidos a la evaluación, y entre sus programas de acción para ayudar a los países a lograr las metas propuestas se encuentran también el refuerzo y la cooperación en las acciones evaluadoras.

El avance de este ambicioso proyecto exige, entre otras cosas, un potente y riguroso sistema de evaluación que permita realizar el seguimiento de las políticas de los distintos países. Algunas de las medidas que será preciso impulsar en el marco del proyecto Metas Educativas 2021 apuntan a reforzar las unidades de obtención de información de los Ministerios de Educación; a ampliar y extender un sistema de indicadores conjunto; a desarrollar entre todos modelos integrales de evaluación de los sistemas educativos y de las escuelas; a elaborar y difundir modelos de evaluación del profesorado; y a apoyar las evaluaciones internacionales que se están llevando a cabo en la región.

Hay que señalar, finalmente, que el proyecto de las Metas Educativas 2021 incluye un sistema de seguimiento y evaluación para valorar su grado de consecución. Así lo entendieron los ministros de Educación cuando, en su declaración final en la Conferencia de El Salvador en mayo de 2008, acordaron también progresar en la elaboración de mecanismos de evaluación regional. La propuesta sobre la que se está trabajando establece informes periódicos de avance, que muestren el nivel de logro alcanzado en las diversas metas, informes específicos, bien temáticos o bien subregionales, e informes conjuntos que pongan en relación los progresos registrados en las Metas Educativas 2021 con otros proyectos internacionales. Una tarea y un esfuerzo compartidos que deben reposar sobre la confianza mutua y la transparencia, basados en el conocimiento científico y práctico acumulado en la región, a partir del esfuerzo que ya se está realizando, y adaptados a la diversidad de puntos de partida, de situaciones y de prioridades educativas de cada uno de los países.

Introducción

Elena Martín y Felipe Martínez Rizo

Comprobar en qué medida se están consiguiendo los objetivos que se persiguen es una actividad que los humanos hacemos muchas veces a lo largo del día. Evaluamos así si vamos o no por buen camino, y es la única forma de mejorar nuestro comportamiento. La evaluación educativa responde a esta misma lógica. Los sistemas escolares tienen que dotarse de procedimientos que les permitan conocer en qué grado se van logrando las metas educativas que la sociedad les ha encargado y, sobre todo, las razones que se vinculan al grado de éxito. Parece legítimo, por tanto, afirmar que la evaluación es un factor nuclear e irrenunciable de la calidad de la enseñanza.

Este libro pretende presentar una visión global de los distintos niveles, ámbitos y dimensiones que se incluyen dentro del amplio campo de la evaluación educativa. A pesar de este objetivo de globalidad, obviamente ha sido necesario acotar el enfoque y el contenido de la obra. Un primer foco se refiere al ámbito geográfico que abarca el libro. La reflexión se centra en el contexto iberoamericano. Se trata de analizar los modelos y actividades que tienen mayor presencia en la región, lo que no quita, desde luego, que se presenten en algunos casos experiencias que se realizan en otros países.

Los capítulos abordan la evaluación en las etapas educativas anteriores a la universidad y otras formas de enseñanza superior. En muchos países iberoamericanos la educación básica sigue presentando grandes retos, y se ha considerado prioritario dedicar este primer libro a su mejora a través de la evaluación. En sucesivos volúmenes de la serie se tratarán otros niveles educativos y otros temas sin duda relevantes.

La voluntad de los coordinadores y autores del libro ha sido dirigirnos al conjunto de los profesionales que trabajan en el sistema educativo y no únicamente a los expertos en el tema. Ello supone intentar mantener un difícil equilibrio entre el rigor y la accesibilidad. Confiamos en haberlo conseguido y remitimos al lector que desee profundizar en algún tema a las referencias bibliográficas que se recogen en las páginas finales.

En el libro se ha buscado conseguir un segundo equilibrio, no menos difícil, entre la valoración positiva de los aciertos que se reconocen en las prácticas evaluativas que se han desarrollado y el análisis de los riesgos y las insuficiencias que se identifican. Para quienes han participado en el libro resulta obvio que la evaluación es necesaria, pero con igual claridad consideramos que contribuir a su mejora implica ser los primeros en seguir reflexionando sobre las dificultades para encontrar la forma de superarlas.

El volumen se organiza en tres partes. La primera de ellas, *La evaluación educativa hoy*, comienza con un capítulo escrito por Alejandro Tiana, que presenta el contexto en el que se inscribe la creciente importancia de la evaluación en el mundo actual, en contraste con lo que ocurría hace poco tiempo, y expone las tensiones que están presentes en el momento de definir las metas de la evaluación, los procedimientos para llevarla a cabo y el uso que de sus resultados se hace.

El siguiente capítulo propone un marco de referencia en el que se puedan situar los diversos acercamientos que se presentarán en la segunda parte, dado que la evaluación no se reduce al acercamiento más conocido, basado en la aplicación de pruebas de rendimiento. La voluntad de cubrir diversos acercamientos a la evaluación en forma amplia y global exigía hacer explícita la perspectiva en que se sitúa el libro, dado que existen opciones muy diferentes, cuando no contradictorias. Esta toma de posición constituye el contenido del estudio debido a la pluma de Felipe Martínez Rizo.

Los dos capítulos restantes de la primera parte ofrecen vistas panorámicas de las evaluaciones que se llevan a cabo en los países iberoamericanos. En el tercer trabajo, Enrique Roca repasa el ámbito de las evaluaciones realizadas por los principales organismos internacionales: IEA, OCDE y UNESCO. Por su parte, Guillermo Ferrer aborda a continuación las evaluaciones que se llevan a cabo en Latinoamérica, tanto en el nivel nacional como en el internacional.

La segunda parte, *Las múltiples caras de la evaluación,* analiza tipos específicos de evaluación. Tres estudios presentan acercamientos que se refieren al conjunto de los sistemas educativos.

El primero de ellos, de Margarita Poggi, analiza un instrumento esencial en la evaluación de la calidad educativa: los sistemas de indicadores, destacando su relación con los principales desafíos que enfrentan los sistemas educativos.

Marta Kisilevsky desentraña una parcela poco estudiada: la evaluación de las políticas educativas, que sin duda requieren de un proceso de reflexión sobre su grado de éxito o fracaso tanto o más que otros elementos del sistema.

Por su parte, Elena Martín se refiere a las pruebas estandarizadas, que evalúan el aprendizaje en la escala de los sistemas educativos, analizando en qué medida ayudan a iluminar las competencias básicas del currículo o si, por el contrario, simplifican las intenciones educativas, al evaluar no tanto lo que se debe, cuanto lo que se puede.

Los tres capítulos restantes de esta parte presentan acercamientos de evaluación que se refieren a ámbitos más restringidos.

El texto de Leda Badilla se centra en la evaluación del centro escolar, como unidad esencial de la calidad de la educación y por ello también de la evaluación. La evaluación de los establecimientos educativos con una finalidad de mejora y no de control es un enfoque poco desarrollado, en especial en América Latina, pese a que es el punto de partida para los siempre reclamados planes de mejora de la calidad.

La evaluación de los docentes es, en este momento, uno de los retos de todos los sistemas educativos: cómo hacerla, de qué forma vincularla al desarrollo profesional, su dependencia o independencia de la evaluación del conjunto del centro escolar. Pedro Ravela revisa desde una lúcida crítica los modelos que se utilizan y propone en su artículo ciertos principios que permitirían impulsar experiencias más positivas.

En el siguiente capítulo se aborda nuevamente la evaluación del aprendizaje, pero ahora desde la perspectiva del aula. Juan Manuel Esquivel revisa los procedimientos de evaluación desde su capacidad para evaluar el rendimiento de los estudiantes de acuerdo con las competencias que se recogen en el currículo como las metas de la educación escolar.

Como pretende indicar su nombre, *El sentido de la evaluación*, la tercera parte del libro parte de la idea central de que la evaluación no es un fin en sí misma, sino que cobra sentido en la medida en que sus resultados son conocidos y aprovechados para sustentar decisiones que lleven a la mejora de la calidad, que es el propósito último que toda evaluación debe perseguir.

El primer capítulo de este bloque se ocupa de la difusión de resultados, una de las grandes dificultades del tema. La complejidad de toda evaluación relevante implica una complejidad similar en la interpretación de sus resultados. Con demasiada frecuencia se difunden conclusiones reduccionistas y simplificadoras, cuando no directamente erróneas. Afrontar este problema supone planificar la difusión, como bien expone Patricia McLauchlan de Arregui.

Finalmente, el texto de Álvaro Marchesi analiza la razón de ser de la evaluación, que no es otra que guiar las políticas educativas. Asumir esta función clave no supone atribuir a la evaluación la exclusividad de las fuentes de las que puede y debe nutrirse la toma de decisiones de los responsables de las administraciones educativas. Encontrar el punto de equilibrio en el uso de los resultados de la evaluación no es una tarea sencilla, pero sin duda es esencial avanzar hacia ello.

Como se ha dicho, en sucesivos volúmenes de la serie que se inaugura con esta obra se tratarán otros temas relevantes para quienes se interesan en la educación. Esperamos que el contenido de los capítulos siguientes enriquezca la visión de sus lectores sobre un tema tan relevante como es el de la evaluación.

La evaluación educativa hoy

Evaluación y cambio educativo: los debates actuales sobre las ventajas y los riesgos de la evaluación

Alejandro Tiana

Una de las personas que ha reflexionado de manera más incisiva sobre el papel que desempeña la evaluación en nuestra sociedad actual, el profesor Ernst House, iniciaba una de sus obras más influyentes, escrita a comienzos de la década de los noventa, hablando de los cambios que él mismo había observado en este campo a lo largo de su vida profesional. Sus palabras, que no me resisto a citar, aunque ya lo haya hecho en alguna ocasión anterior, resultan muy reveladoras: "Cuando comencé mi carrera en evaluación hace más de veinticinco años, reuní todos los trabajos que pude encontrar en una pequeña caja de cartón en un rincón de mi despacho y los leí en un mes. Desde entonces, la evaluación ha pasado de ser una actividad marginal desarrollada a tiempo parcial por académicos a convertirse en una pequeña industria profesionalizada, con sus propias revistas, premios, reuniones, organizaciones y estándares" (House, 1993, p. 1, traducción del inglés).

Si en 1993 House hablaba de una "pequeña industria profesionalizada", hoy creo que se puede hablar de una industria en toda regla. En efecto, al echar la vista atrás a lo que ha sucedido en las dos últimas décadas, observamos que por todas partes se han establecido sistemas de evaluación y de acreditación tanto en el ámbito empresarial como en los servicios públicos (no solo los educativos), que se han constituido sociedades nacionales, regionales e internacionales de evaluación, que se han puesto en marcha ambiciosos proyectos de elaboración y difusión de indicadores en campos muy diversos, que se ha consolidado la oferta de formación académica en materia de evaluación: en suma, que la evaluación ha encontrado su lugar propio en el campo del saber y en el de la actuación pública y privada.

Al contemplar lo que ha sucedido en el mundo específicamente educativo, no podemos sino reforzar esa apreciación, pues desde finales de la década de los ochenta, y sobre todo en los últimos quince años, se ha generalizado la constitución de sistemas nacionales de evaluación educativa, que desarrollan distintos tipos de actividades. La aplicación de pruebas estandarizadas de rendimiento, la construcción de indicadores, la evaluación y autoevaluación de los centros educativos o la evaluación del desempeño de los docentes no son sino una pequeña muestra del tipo de iniciativas puestas hoy en marcha en la inmensa mayoría de los países iberoamericanos. Hoy evaluamos (o, al menos, hablamos de evaluar) todo el ámbito de la educación: el rendimiento de los alumnos, los currículos escolares, los centros educativos, los programas específicos o sectoriales y el trabajo de los docentes.

Como consecuencia de ese proceso, la evaluación ha ido abarcando ámbitos progresivamente más amplios, al tiempo que se ha diversificado, reforzando así una tendencia que ya se apuntaba hace

más de diez años (Tiana, 1996). Además, las incipientes iniciativas internacionales de evaluación educativa de la segunda mitad del siglo xx han dado paso a operaciones amplias, complejas y de gran impacto político y mediático, como el proyecto PISA, desarrollado por la Organización para la Cooperación y el Desarrollo Económico (OCDE) o los estudios TIMSS y PIRLS, promovidos por la Internacional Association for the Evaluation of Educational Achievement (IEA). Puede decirse que, frente a quienes consideraban que se trataba simplemente de una moda, que bien podría ser pasajera, la evaluación ha llegado para quedarse en nuestros sistemas educativos, al menos por un largo tiempo.

Siendo así la situación, merece la pena reflexionar seriamente acerca de cuál es la función que hoy día desempeña la evaluación en el ámbito educativo. Y una de las conexiones más importantes que cabe establecer es la que relaciona la evaluación con el cambio en educación. La reflexión sobre ese asunto es el objeto de estas páginas.

EDUCACIÓN Y CAMBIO

Si hay una realidad indisociable de la tarea educativa, creo que es precisamente el cambio, la transformación. La esencia última de la educación consiste en producir cambios en los más jóvenes, para convertirlos en personas adultas capaces de afrontar con garantías de éxito los desafíos que se les plantearán a lo largo de la vida. Por lo tanto, educar consiste en última instancia en producir y acompañar procesos de cambio personal. Pero la meta a la que ese proceso se orienta, esto es, una vida adulta provechosa y satisfactoria en una sociedad determinada, es asimismo cambiante, pues las sociedades han cambiado siempre a lo largo de la historia y continúan haciéndolo, con un ritmo además cada vez más rápido. Por lo tanto, no se puede concebir una educación estática (aunque algunos parecieran quererla en ocasiones), sino que debemos aprender a convivir con el cambio y a aprovecharlo para lograr nuestros propósitos.

Ahora bien, el cambio es una realidad ambivalente. Por una parte, es un fenómeno natural, que se produce al margen de nuestra voluntad. Querámoslo o no, el mundo cambia en todos sus aspectos. Es cierto que no hay que considerar que la transformación represente siempre un progreso o avance, pues por desgracia también puede suponer lo contrario, pero es una quimera soñar con una realidad estática. Por otra parte, pretendemos ser capaces de orientar el cambio para que nos permita mejorar, avanzar, también en educación. No nos resignamos a que el cambio se produzca de forma ciega, sin nuestra intervención, sino que nos planteamos incidir en él. Por eso diseñamos reformas y programas de mejora educativa, por ejemplo.

Tratándose de una realidad ambivalente, no es extraño que haya atraído mucha atención. El estudio del cambio en educación cuenta ya con una larga tradición y las investigaciones realizadas ponen de manifiesto su complejidad. Algunos autores han dedicado a esa tarea una buena parte de sus esfuerzos académicos, realizando aportaciones de gran interés (Fullan, 1993 y 1999). Los trabajos llevados a cabo han puesto de manifiesto la complejidad de los procesos de cambio y nos han alertado acerca de los riesgos que corremos al interpretar dichos fenómenos de manera unilateral o al creer que nuestros propósitos son directamente trasplantables a la realidad, sin tener en cuenta los condicionantes a que se enfrentan. Por eso, siempre que hablamos de cambio debemos ser extremadamente cautelosos. Por ejemplo, algunos de los problemas causados por los procesos recientes de reforma educativa emprendidos en diversos lugares derivan del olvido de esa complejidad (Saranson, 2003).

LA EXPANSIÓN DE LA EVALUACIÓN COMO RESULTADO DEL CAMBIO DE LOS SISTEMAS EDUCATIVOS

Al analizar las relaciones que pueden establecerse entre la evaluación y el cambio educativo, lo primero que debemos destacar es que el proceso de auge y expansión de aquella es en buena medida consecuencia de este. En efecto, no se entenderá correctamente el papel y la función que desempeña hoy día la evaluación si la desvinculamos de los cambios experimentados por los sistemas educativos actuales.

No cabe duda de que los sistemas educativos están hoy sometidos a fuertes tensiones e incertidumbres. Hay quien habla de crisis y quien prefiere hacerlo, con palabras más medidas, de transformación profunda. Sin querer entrar aquí en ese debate conceptual y terminológico, lo cierto es que el ritmo actual de cambio de los sistemas educativos es probablemente el más intenso desde su constitución, que recordemos que se produjo en paralelo con la formación del Estado nacional y el tránsito de una sociedad estamental a una sociedad de clases, a finales del siglo XVIII y comienzos del XIX.

Son varios los factores que han provocado esa situación de incertidumbre. Entre ellos hay que destacar, por su importancia, la puesta en cuestión del Estado-nación por el auge de la globalización y de los procesos de articulación y convergencia supranacional; la extensión del concepto del aprendizaje a lo largo de la vida, que ha obligado a relativizar el papel de la educación formal; la renovada conciencia de la estrecha interrelación entre la educación y el desarrollo económico y social, que ha vuelto a atraer la atención hacia los fenómenos educativos, y la gran influencia que ha ejercido la ideología neoliberal, que ha puesto en cuestión el papel que debe desempeñar el Estado en la prestación de los servicios públicos.

Flexibilidad, descentralización y autonomía escolar

Como resultado de la confluencia de todos estos factores, los sistemas educativos afrontan nuevas demandas y se han visto obligados a desarrollar nuevos modelos de actuación. Sobre todo, se les ha exigido que acentúen su flexibilidad. Si el objetivo actual consiste en proporcionar a todos los jóvenes, sin distinción de sexo, etnia u origen socioeconómico, una educación común, equivalente, durante un largo período de tiempo (generalmente desde los seis años hasta los dieciséis), resulta ineludible hacer hincapié paralelamente en la atención a la diversidad. En efecto, una educación que pretende ser común no puede concebirse de manera uniforme, ni impartirse de forma idéntica para todos, sino que debe adaptarse a la diversidad de situaciones, condiciones de partida, intereses y contextos. Por lo tanto, las escuelas deben actuar en consonancia con esa diversidad, adaptando su organización, su currículo y su modo de funcionamiento a las circunstancias cambiantes en que se desenvuelven, esto es, aplicando el criterio de flexibilidad. Es lo que ha dado en denominarse una nueva personalización de la acción educativa (OCDE, 2006).

La demanda de flexibilidad ha repercutido directamente en la descentralización de los servicios educativos, de un lado, y en el refuerzo de la autonomía escolar, de otro. No hay que confundir ambos fenómenos, pues aunque estén generalmente relacionados son claramente distinguibles. Así, existen Estados que han descentralizado notablemente la educación hacia las regiones, comunidades o municipios y, sin embargo, continúan sometiendo a sus escuelas a una regulación bastante rígida, como también se dan casos a la inversa. Considerada la situación en conjunto, es

evidente que la última década ha asistido a un intenso movimiento descentralizador en el ámbito internacional, que también se ha dejado sentir en los países iberoamericanos. El cambio así producido no ha dejado de suscitar debates, pues en ocasiones la descentralización se ha entendido como un traspaso del sistema escolar por parte del Estado hacia entidades territoriales menores, sin ir siempre acompañado de los recursos suficientes y sin insistir en la regulación estatal que debiera asegurar la prestación del servicio público de la educación en condiciones de equidad. Pero, en términos generales, estos procesos se han situado en el contexto de una tendencia más amplia que se ha extendido a escala planetaria. En contraposición con lo que ha ocurrido con la descentralización, el refuerzo de la autonomía escolar, aunque ha estado muy presente en el discurso educativo, no se ha visto siempre acompañado de medidas que lo hiciesen posible. Por lo tanto, hay que señalar que ambos fenómenos han sido asimétricos y no han tenido necesariamente una evolución paralela.

El énfasis en la descentralización y en la autonomía escolar ha tenido un efecto evidente sobre la expansión de la evaluación educativa, como consecuencia del cambio que ha introducido en los modos de control de la educación. En los sistemas educativos uniformes y centralizados, el control se basa en una combinación de tres factores: una producción normativa generalmente reglamentista y detallada, que fija con precisión el marco de actuación de las escuelas y del profesorado; una regulación minuciosa de los procesos escolares, sobre todo en relación con el currículo que se debe desarrollar, los manuales que hay que utilizar, los tiempos escolares que pautan la vida escolar y las agrupaciones del alumnado, y la actuación de los servicios de inspección y supervisión educativa que aseguran el cumplimiento de las normas establecidas.

Por el contrario, en los sistemas más descentralizados o con mayor autonomía escolar, esos procedimientos de control, sin perder totalmente su funcionalidad, han sido total o parcialmente sustituidos por otros, basados fundamentalmente en la evaluación y el seguimiento de los resultados obtenidos. A cambio de la relajación del detalle de la normativa general y de la menor regulación de los procesos, las administraciones han pasado a controlar más de cerca el rendimiento del sistema y de sus componentes. Es cierto que la medición y la valoración de los resultados que obtienen los alumnos y las escuelas son tareas complejas y que no están exentas de complicaciones, como se plantea en otros capítulos de este mismo libro, pero ello no es excusa para no abordarlas. Así pues, las autoridades educativas han comenzado a percibir la evaluación como un instrumento útil para la administración y el control de la educación, y en consecuencia la han fomentado. No hay que interpretar que el proceso haya sido plenamente coherente o similar en todos los países, ni tampoco que se haya completado en todas partes, pero, en términos generales, creo que la descripción anterior refleja adecuadamente el cambio que se ha ido produciendo.

El lector reconocerá este tipo de argumentación y su lógica subyacente, que enlaza flexibilidad con descentralización, autonomía y evaluación, en la puesta en marcha de muchos sistemas nacionales de evaluación educativa durante las dos últimas décadas. Este cambio ha provocado diversos debates, relativos, por ejemplo, a la legitimidad de la descentralización y de la evaluación como un nuevo mecanismo de control, a la complejidad de la evaluación de los resultados escolares, o a la articulación entre currículo y evaluación, especialmente en lo que se refiere al tratamiento que reciben las áreas menos susceptibles de medición estandarizada. Las discusiones han sido en ocasiones intensas y cargadas de sentido. Pero la conclusión, cuando echamos la vista al proceso desarrollado, es que no han frenado la expansión de los sistemas nacionales de

evaluación, seguramente porque están bien adaptados a las nuevas demandas que se plantean a los sistemas educativos.

Devolución de responsabilidades y rendición de cuentas

Los sistemas educativos no solo se enfrentan a una demanda de mayor flexibilidad en su modo de organización y en su funcionamiento, sino que también se les reclama que promuevan una mayor implicación de los ciudadanos y de la llamada sociedad civil en la gestión de los asuntos educativos. En última instancia, se trata de dos fenómenos relacionados, de dos caras de la misma moneda, si se quiere. La descentralización y la autonomía escolar han sido frecuentemente interpretadas como parte de un movimiento político de devolución de las responsabilidades educativas a instancias de gestión más cercanas a la ciudadanía.

Este planteamiento tiene profundas connotaciones políticas y ha generado importantes debates. Por una parte, se acepta generalmente que la gestión de los asuntos educativos debe incorporar a todos los sectores que componen la comunidad educativa. Ese es el fundamento de que los sistemas educativos actuales hayan incorporado el principio de participación democrática en su funcionamiento, aunque lo lleven a la práctica de maneras diversas. Aún más, por lo general nos planteamos que no basta con asegurar una participación formal, al estilo de la que se produce en los denominados consejos escolares, sino que se debe dar un paso más allá y lograr una efectiva implicación de la comunidad en los asuntos educativos. Ese es el motivo de que, por ejemplo, muchos de los sistemas nacionales de evaluación estudien hoy el grado de implicación de las familias en la educación de sus hijos. Pero, por otra parte, la citada devolución de responsabilidades suscita algunos recelos, como los derivados del rechazo a los planteamientos neoliberales que defienden que el Estado debe achicarse y actuar de manera subsidiaria a la actuación de familias y comunidades, sembrando así la semilla de nuevas desigualdades en materia de educación. Para quienes compartimos dichas críticas, la devolución de responsabilidades tiene justificación social y política siempre que no implique el descuido del papel crucial del Estado en la prestación del servicio público educativo o el abandono a las simples fuerzas del mercado de un ámbito tan fundamental de actuación pública como es la educación.

Entre las principales consecuencias que ha producido este complejo movimiento de devolución de responsabilidades, descentralización educativa y refuerzo de la autonomía escolar cabe destacar la aparición de una nueva demanda de rendición de cuentas en el ámbito educativo. Los sistemas educativos reciben importantes recursos y se espera de ellos que realicen una contribución relevante al desarrollo económico y social. Las escuelas gozan de la confianza pública para contribuir a formar lo mejor posible a los ciudadanos más jóvenes. A cambio, los sistemas educativos y las escuelas deben dar cuenta de la utilización que hacen de los recursos que se les confían y, sobre todo, de los resultados que consiguen. Frente a la imagen muy difundida del sistema educativo como una caja negra, de la que no se conoce qué sucede exactamente en su interior, se extiende la demanda de una mayor transparencia y una rendición de cuentas coherente con los principios de la gestión democrática.

La difusión de ese planteamiento de rendición de cuentas ha contribuido notablemente a la expansión de los sistemas nacionales de evaluación en la última década. Los estudios que llevan a cabo pretenden dar respuesta a las preguntas que se formulan acerca de los logros educativos conseguidos, los avances experimentados, el grado de equidad alcanzado o el uso que se hace de los recursos existentes, estableciendo además comparaciones entre países, regiones e incluso es-

cuelas. Las respuestas dadas a esos interrogantes son consideradas clave para adoptar estrategias de actuación, para establecer prioridades políticas, para incentivar a las escuelas y al profesorado, para mostrar a las familias y a los ciudadanos lo que se ha conseguido (y lo que falta por conseguir) y para objetivar el debate público acerca de la educación.

LA EVALUACIÓN COMO INSTRUMENTO DE CONDUCCIÓN DE LOS PROCESOS DE CAMBIO

Pero aunque el auge y la expansión de la evaluación se expliquen en buena medida como consecuencia de las transformaciones experimentadas por los propios sistemas educativos, no hay que creer que se trate de procesos ciegos o faltos de dirección. Como se indicaba más arriba, la pretensión de controlar y dirigir el cambio es una constante en la educación. Y la relevancia adquirida por la evaluación también se explica parcialmente por los objetivos que se le han asignado en los sistemas educativos actuales, dicho de otro modo, por lo que se ha querido conseguir mediante su aplicación generalizada.

Los nuevos modos de gestión educativa a que antes se hacía referencia han hecho aparecer y difundirse nuevos modelos de administración. Uno de los que más han llegado a calar es el modelo de conducción o seguimiento. Su imagen sería la de un sistema que funciona y se desarrolla con notable autonomía interna (el sistema educativo) y al que interesa ir orientando (desde la administración) para asegurar que se alcanzan los objetivos establecidos. En ese modelo, la evaluación desempeñaría un papel fundamental para controlar la dirección y el ritmo del cambio y para comprobar la distancia que aún resta para alcanzar los objetivos marcados.

El modelo de conducción o seguimiento era difundido a comienzos de los noventa por Gilbert de Landsheere, quien lo definía como "la toma de decisiones al nivel macroscópico (es decir, de un sistema o subsistema educativo) o microscópico (es decir, de un establecimiento o incluso una clase), sobre la base de constataciones tan objetivas como sea posible, relativas al estado, al funcionamiento o a los productos del sistema" (De Landsheere, 1994, p. 8, traducción del francés). En tiempos recientes se ha planteado un nuevo debate académico y político acerca de la posibilidad de realizar una conducción basada en los resultados de la educación, que viene a ser una actualización de esas mismas ideas y que entiende que se trata del modelo mejor adaptado a las circunstancias actuales.

Así considerada, la evaluación sería uno de los instrumentos fundamentales para pilotar el cambio. Como decían los ministros iberoamericanos de Educación, reunidos en Buenos Aires en 1994, "bien podría afirmarse que la evaluación no es sino una forma distinta y nueva de hacer política y, más concretamente, de explicar las políticas por medio de los logros alcanzados o, eventualmente, fallidos" (OEI, 1994, p. 5).

Estas últimas palabras ponen el acento en una observación fundamental para situar la evaluación en su justo lugar y que no nos cansaremos de repetir: la evaluación no sustituye a la política educativa, sino que es un instrumento a su servicio. Es importante, eso sí, pero no deja de ser un simple instrumento. Dicho con otras palabras, evaluar no mejora la realidad que se evalúa, lo que consigue es proporcionar una información tan rigurosa y detallada como queramos para conocerla mejor y actuar en consecuencia. A partir del diagnóstico que hagamos y del juicio que emitamos sobre el aspecto o aspectos de la realidad que evaluamos, podremos adoptar medidas que permitan superar los puntos críticos detectados y adoptar decisiones orientadas a la mejora.

Por seguir con el símil de la conducción, la calidad de los instrumentos utilizados no sustituye la pericia del piloto, que es quien debe adoptar las decisiones que convengan para llegar al punto de destino.

LAS POSIBILIDADES Y LOS RIESGOS DE LA EVALUACIÓN

Para finalizar estas reflexiones me detendré a analizar con algo más de detalle algunas cuestiones que han ido apareciendo de manera lateral o indirecta en las páginas anteriores. El lector habrá ido encontrando referencias realizadas más o menos de pasada a algunos debates actualmente vigentes sobre asuntos tales como las posibilidades y las limitaciones de la evaluación de los resultados de los estudiantes y de las escuelas, los usos posibles de la evaluación en sistemas descentralizados, la demanda de participación e implicación de los agentes educativos o la siempre complicada conexión entre evaluación, cambio y mejora. Esta relación no constituye un catálogo completo de las cuestiones que hoy preocupan en este ámbito, sino que debe entenderse como una simple muestra de las controversias suscitadas.

Y es que la evaluación es una actividad compleja y sometida a discusión. No debe extrañar que así sea, puesto que en última instancia implica la emisión de juicios de valor sobre los objetos evaluados. Por una parte, plantea interrogantes acerca del rigor y la exactitud de la información recogida, la calidad de los análisis realizados o la objetividad de las conclusiones alcanzadas. Por otra parte, la propia situación de valoración, consustancial con la evaluación, obliga a plantearse asuntos tales como la participación de las audiencias interesadas en la evaluación, la transparencia de sus procedimientos o la justicia de los juicios emitidos.

Evaluación y mejora: la utilidad de la evaluación

Un primer asunto que suscita posiciones encontradas se refiere a la relación existente entre evaluación y mejora. Algunos evaluadores consideran que su función consiste básicamente en examinar cuidadosamente la realidad que analizan, aplicar métodos y técnicas apropiados para profundizar en ella y ofrecer una descripción y una valoración ajustadas de su situación y estado. Para quienes así piensan, ese es el núcleo de su tarea y ahí acaba su función.

Para otro grupo de evaluadores, en cambio, su actuación no puede desligarse del objetivo último de mejorar la realidad que evalúan. Consideran que la evaluación supone una intervención enérgica, e incluso en ocasiones algo agresiva, ya que busca qué hay debajo de la primera apariencia de las cosas, se pregunta acerca de cuáles son las dinámicas establecidas y la posición de los distintos actores y emite juicios de valor sobre el funcionamiento o los logros de un programa, una institución o un sistema. Por consiguiente, consideran que la contrapartida a esa intervención consiste en comprometerse con la mejora. Obviamente, lo normal (y lo más razonable) es que no le corresponda al evaluador tomar las decisiones pertinentes tras la evaluación, pero es distinto desarrollar la tarea evaluadora poniendo el final del proceso en la entrega del informe de evaluación o en la interacción con quienes deben decidir, por ejemplo.

La posición adoptada al respecto por el evaluador depende en ocasiones del tipo de evaluación que se aborde. No es lo mismo evaluar un programa educativo o una institución escolar que valorar el rendimiento global de los estudiantes de una región o el funcionamiento de un sistema educativo. Mientras que en unos casos la implicación es muy cercana y directa con los actores educativos, en otros es una relación mediada, generalmente a través de los administradores o los responsables

políticos. Pero, en uno y otro caso, son muchos los evaluadores que consideran que la evaluación debe tener en cuenta la vertiente de la mejora si no quiere quedarse en un simple, aunque sea valioso e interesante, ejercicio analítico.

Es necesario, no obstante, clarificar esta posición, dejando claro que con ella no se trata de defender la participación directa y explícita de los evaluadores en la toma de decisiones. Por muchos motivos conviene separar ambos procesos y distinguir nítidamente a los responsables de una y otra actuación, pero ello no quiere decir que el evaluador se desentienda de la interpretación que se hace de sus resultados o de las aplicaciones que se realizan de la evaluación para influir sobre la realidad.

Desde este punto de vista, conviene también subrayar que la evaluación no debe pretender dar orientaciones precisas para la toma de decisiones. De hecho, algunos evaluadores se lamentan de que sus conclusiones no son suficientemente tenidas en cuenta a esos efectos. Se trata de una queja derivada de un equívoco, que consiste en pensar que la evaluación debe servir de guía de actuación o, como se dice hoy, de hoja de ruta para el cambio. No obstante, los evaluadores tienden más bien en la actualidad a defender el uso iluminativo de la evaluación. De acuerdo con la misma, la evaluación no tendría como función única o principal la de ofrecer elementos para la toma de decisiones, sino que también cumpliría un papel importante contribuyendo a mejorar el conocimiento de los procesos educativos y arrojando luz sobre los mismos.

Objetividad, credibilidad y justicia de la evaluación

Si descendemos a la tarea de evaluación propiamente dicha, surge una segunda cuestión que también suscita controversia. Se trata de la importancia que se debe conceder a la objetividad de la evaluación y las consecuencias que ello tiene para su credibilidad y para la justicia que conlleva su actuación.

La objetividad es una aspiración irrenunciable de cualquier estudio con pretensiones científicas, pero se sabe bien la dificultad que implica, sobre todo cuando su objetivo consiste en emitir una valoración. La valoración y el juicio, que constituyen precisamente el núcleo de la acción evaluadora, implican la referencia a valores y, por lo tanto, dificultan una percepción puramente objetiva. En consecuencia, toda acción evaluadora está forzosamente sometida a una tensión interna entre conocimiento e interpretación, que no siempre resulta fácil de resolver y en la que se juega su crédito.

Pese al reconocimiento de tales limitaciones y de la dificultad de conciliar rigor, objetividad y valoración, el juicio que emitan la sociedad y los ciudadanos acerca de la credibilidad de los datos obtenidos y de las interpretaciones realizadas constituye la clave del éxito o fracaso de las políticas de evaluación. La búsqueda y consecución de tal credibilidad se convierte, por lo tanto, en uno de sus objetivos fundamentales.

La cuestión de la credibilidad lleva normalmente aparejada la de la independencia institucional de los sistemas de evaluación o, dicho en otros términos, la de su imparcialidad. Obviamente, los modos de organización institucional de la tarea de evaluación del sistema educativo variarán considerablemente en función de las circunstancias concretas en que se desarrollen, sin que puedan determinarse reglas universales. Los organismos encargados de realizar dicha tarea pueden vincularse más o menos, de un modo u otro, a los distintos poderes del Estado y a los diversos escalones

administrativos, pero su percepción como instituciones con un amplio margen de independencia, conseguida muchas veces a través del equilibrio interno de poderes, es la principal garantía de éxito en su tarea y de reducción de los riesgos que plantea.

En este contexto de refuerzo de la credibilidad por la vía del equilibrio de poderes, conviene destacar que uno de los riesgos de la evaluación sobre los que se ha alertado consiste en atender solamente a los requerimientos de quienes la encargan o financian, olvidando las posiciones de otros actores. Para evitarlo y para reforzar las posibilidades de la evaluación se ha introducido el concepto de "audiencia", denominación con la que se hace referencia a todas las personas y grupos afectados o que tienen intereses legítimos en un proceso determinado de evaluación.

Íntimamente ligada a esta cuestión aparece también la del principio de justicia que debe regir la actuación evaluadora. Es un asunto que se pone de manifiesto, por ejemplo, cuando se realizan evaluaciones que tienen consecuencias directas para las personas o las instituciones. Uno de los casos más claros es el de la evaluación de centros escolares con la intención de adoptar decisiones sobre ellos o de establecer clasificaciones de cualquier tipo. Tanto los agentes implicados en la evaluación como sus destinatarios deben preguntarse si la evaluación refleja de manera justa la situación de las escuelas y actuar de forma que lo garantice. Por ejemplo, no se pueden pedir los mismos resultados a escuelas que atienden a un alumnado de distinto potencial o diferente origen social y cultural. Parece más lógico compararlas introduciendo algunos índices correctores de su situación. No otra cosa han pretendido hacer los distintos modelos de "valor añadido" en educación que se han desarrollado en los últimos años y que tienden a aplicarse cada vez más frecuentemente. El riesgo de olvidar o ignorar estas diferencias amenaza seriamente las posibilidades de la evaluación.

Coherencia de la evaluación con las políticas educativas

Una última cuestión que merece la pena abordar aquí y que también ha generado debate tiene que ver con la coherencia que debe existir entre la política de evaluación y otras políticas educativas. Se trata de fomentar los efectos beneficiosos de la evaluación para el desarrollo de la educación y de evitar sus efectos perversos. La experiencia nos demuestra que la aparición de efectos no deseados ni deseables en este ámbito es más frecuente de lo que debiera, lo que obliga a no ignorarlos y a prestarles atención.

Hay que insistir reiteradamente en que las políticas de evaluación del sistema educativo no deben concebirse al margen de las políticas generales de desarrollo del mismo. En última instancia, las primeras deben colaborar al logro de los objetivos de la educación. Así, por ejemplo, aunque pocos se opondrán a la utilización de la evaluación para la mejora de la eficacia y la eficiencia del sistema educativo, existen divergencias importantes entre la aplicación práctica de dicho planteamiento en un modelo político de corte neoliberal o preocupado prioritariamente por la excelencia (interesado, por ejemplo, en fomentar la competencia entre escuelas y redes escolares, presuponiendo que es una de las claves para la mejora) y otro de corte socialdemócrata o preocupado primordialmente por la equidad. Por ello, las políticas de evaluación no deben concebirse como una parcela aislada dentro del amplio campo de la política y la administración educativas.

Otro tanto podría decirse de la conexión existente entre evaluación y currículo. Este último campo ha sido objeto de especial atención en las reformas educativas de la década de los noventa, aunque hoy tendamos a verlo como un elemento que debe integrarse en una visión más sistémica del cambio. En cualquier caso, el currículo es y seguirá siendo un elemento central en los procesos

de enseñanza y aprendizaje. Por ese motivo, existe el riesgo, a veces convertido en realidad, de que las políticas de evaluación no estén completamente de acuerdo con las políticas curriculares o incluso de que se contrapongan a ellas. Por ejemplo, la difusión actual del modelo curricular que conlleva la definición de unas competencias básicas que deben lograr todos los jóvenes tiene el riesgo de contraponerse a un modelo de evaluación del rendimiento basado en un concepto más tradicional del aprendizaje, limitando así sus posibilidades para el cambio.

Para ilustrar estas reflexiones, se pueden recordar los debates que recientemente se han planteado acerca del posible impacto negativo que pueden tener las evaluaciones en gran escala sobre las prácticas docentes, pudiendo llegar incluso a ejercer un efecto perverso. No puede decirse que la aplicación de pruebas estandarizadas de rendimiento tenga necesariamente un efecto negativo como el que a veces se le ha achacado, pero hay que tener presente que existe un riesgo real de que así sea, especialmente cuando los docentes se orientan a preparar a los alumnos para dichas pruebas, descuidando los aspectos del currículo que no quedan cubiertos por ellas. El riesgo de que se produzca ese efecto perverso es aún mayor cuando las pruebas son demasiado simples, cuando sus resultados tengan un alto impacto sobre el profesorado, los estudiantes y las escuelas, y en la medida en que los usuarios no tengan la capacidad o la formación adecuada para entender el alcance real y las limitaciones de ese tipo de evaluaciones. Por ese motivo, el diseño de pruebas de evaluación del rendimiento a gran escala exige un cuidado especial y una consideración detenida de los usos que de ellas se pretende hacer.

A modo de conclusión, debe reconocerse que la evaluación ofrece importantes posibilidades para la mejora de la educación, pero también implica algunos riesgos que no deben obviarse. En estas páginas se ha pretendido analizar la evolución que ha experimentado recientemente la evaluación en el ámbito de la educación, poniéndola en conexión con los cambios que viven los sistemas educativos. Frente a quienes han subestimado la contribución que la evaluación puede hacer y han rebajado la importancia del lugar que ha llegado a ocupar, hay que insistir en las ventajas que encierra. Pero también hay que ser realista y aceptar sus limitaciones, sus riesgos e incluso sus posibles efectos perversos, si se quiere que su desarrollo contribuya efectivamente a la mejora de la educación que reciben nuestros jóvenes, que es el objetivo último que debe orientar nuestra actuación.

La evaluación de la calidad de los sistemas educativos: propuesta de un modelo

Felipe Martínez Rizo

INTRODUCCIÓN

La idea central que orientó la redacción de este capítulo es la de que la calidad de un sistema educativo es un objeto de análisis de tal complejidad, que ningún acercamiento particular puede dar cuenta de ella de manera razonablemente completa; para ello es necesario combinar trabajos de diferente naturaleza que, en conjunto –por ello se habla de un *modelo*–, consigan ese propósito.

La mayor parte de los demás capítulos de la obra se referirán en especial a algunas de las formas de aproximarse a diferentes aspectos de la calidad educativa; en este se ofrece un marco para situar cada acercamiento en un conjunto coherente. Se comienza además con un breve recordatorio sobre el desarrollo de las pruebas de aprendizaje en gran escala, el acercamiento más conocido a la evaluación de la calidad de los sistemas educativos, aunque no el único, como veremos.

El desarrollo de dichas evaluaciones en gran escala, más allá de las individuales de los alumnos, es un fenómeno reciente, lo que no debe extrañar si se tiene en cuenta que los mismos sistemas educativos, en su forma actual, no son tampoco muy antiguos. Desde tiempos remotos, en efecto, las sociedades han buscado formas de apoyar a las familias en el proceso de socialización de las nuevas generaciones, pero solo a partir de los siglos XVIII y XIX surgieron los rasgos de los sistemas educativos modernos: la universalidad y la obligatoriedad; la gratuidad; la laicidad; la organización de los alumnos en grados, y la existencia de docentes profesionales. Un rasgo más que, de alguna manera, engloba a los anteriores, es el carácter público de la educación.

Todavía en el XIX las escuelas de primeras letras no tenían esos rasgos: el acceso a ellas distaba mucho de ser universal y, desde luego, no era obligatorio. Las familias cubrían el costo, puesto que no se trataba de un servicio público. Los maestros no tenían el carácter profesional de los niveles superiores y su organización gremial fue tardía. Los alumnos no se organizaban en grados, sino que formaban un solo grupo, en el que convivían los más adelantados y los de más reciente incorporación. No había una edad prescrita para iniciar el trayecto escolar, ni etapas definidas en él o un número de años para recorrerlo: la enseñanza terminaba cuando, a juicio del maestro, un alumno había alcanzado un nivel suficiente en la lectura, la escritura y la aritmética elemental.

A partir de mediados del XIX, esas escuelas fueron dejando el lugar a los sistemas modernos, de escuelas a las que deben asistir todos los niños de cierta edad, sostenidas con recursos públicos, en las que enseñan maestros profesionales, preparados en instituciones especiales. Para atender a grupos numerosos surgieron modelos tutoriales, como el *lancasteriano*. Luego se extendió el modelo que llegaría a ser universal: la escuela básica organizada en grados, en la que los alumnos

deben comenzar a una edad determinada un trayecto en el que cada año que transcurre implica también, normalmente, el tránsito a un grado superior (Gaulupeau, 1992; Goodlad y Anderson, 1987; Martínez Rizo, 2006a; Oakes y Heckman, 1994).

La organización escolar por grados trajo consigo una nueva necesidad: decidir, al fin de cada grado, si un alumno está en condiciones de pasar al siguiente. En los sistemas tradicionales, el maestro valoraba el avance de los alumnos pero, al haber un solo grupo, no se requería ese tipo de decisiones. El carácter elitista de las escuelas, además, traía consigo una razonable homogeneidad de los alumnos. La masificación de los sistemas educativos trajo consigo creciente heterogeneidad, al llegar a la escuela los hijos de familias acomodadas y los de familias pobres, que no pueden apoyar el trabajo escolar.

En sus primeras épocas, los sistemas educativos modernos no contaban con mecanismos de evaluación distintos de los que utilizaba cada maestro, que se juzgaban suficientes para asegurar que quienes terminaban la escolaridad obligatoria tuvieran un manejo aceptable de la lectura, la escritura y la aritmética. Surgieron mecanismos de control de calidad, con los inspectorados. La expansión de la educación básica hizo cada vez más evidente, sin embargo, que terminar un nivel no garantizaba que se dominaran ciertos conocimientos. Esto llevó al desarrollo de las pruebas en gran escala.

Durante la primera mitad del siglo XX se desarrolló la llamada "teoría clásica de las pruebas"; durante la segunda hubo avances metodológicos diversos, como los modelos de respuesta al reactivo; en las últimas décadas del siglo, en un número creciente de países, el interés por tener mediciones del aprendizaje alcanzado por los estudiantes aumentó y llevó a la extensión del uso de las pruebas de gran escala. Es paradigmático, pero no excepcional, el caso de Estados Unidos, desde el movimiento de *minimum competency testing* de la década de 1980, hasta la Ley *No Child Left Behind* del presidente Bush (Martínez Rizo, 2002). En el capítulo de Enrique Roca en este libro se puede ver lo relativo a la extensión de las evaluaciones internacionales, de finales del siglo XX y principios del XXI.

COMPLEJIDAD DE LA TAREA Y ELEMENTOS PARA ABORDARLA

Poca duda cabe de la necesidad de tener evaluaciones que den una visión objetiva de la calidad de la educación; hoy parecen indispensables para asegurar una gestión eficiente de sistemas sociales tan importantes. Pero hacer evaluaciones adecuadas es una tarea compleja, que debe cuidarse mucho para evitar resultados contraproducentes.

La tarea de evaluar un sistema educativo es difícil por una doble complejidad: una de naturaleza empírica, del objeto de evaluación, y otra de carácter teórico, derivada de las múltiples facetas de las nociones mismas de calidad y evaluación. Reflexionar sobre estas múltiples fuentes de complejidad es un antecedente necesario para entender la propuesta, que se hará luego, de un sistema *integral* de evaluación.

Sobre la complejidad de los sistemas educativos, pensemos que aun en países pequeños deben atender a decenas de miles de alumnos, con miles de profesores, en centenares de escuelas. Además del tamaño, son muchos los aspectos que se deben evaluar y sus circunstancias: hay que evaluar el rendimiento de los alumnos, pero también a maestros, directivos y personal de apoyo, el currículo, los materiales derivados del mismo, los planteles, etc.

Para diseñar evaluaciones que den cuenta adecuadamente de tan grandes sistemas habrá que tener presente su propia complejidad, así como la dificultad teórica, la que se deriva de las dimensiones de la noción de calidad, los diversos propósitos que puede perseguir una evaluación y las características de los acercamientos que se han de utilizar. No es posible hacer evaluaciones que sean al mismo tiempo comprensivas y profundas. Como pasa con los instrumentos ópticos, en los que el campo de visión se reduce a medida que aumenta la aproximación, en evaluación es inevitable el compromiso entre amplitud y detalle.

La noción de calidad educativa y sus dimensiones

Una idea pobre de la *calidad del sistema educativo* pierde de vista aspectos importantes, como ocurre si se reduce la calidad al nivel de aprendizaje que alcanzan los alumnos, o se contraponen cobertura y calidad, como si la una no tuviera que ver con la otra.

Suele decirse: antes prevalecía el interés por la *cantidad*, en referencia a la cobertura; ahora debe preocuparnos la *calidad*. Pero un sistema educativo no podrá considerarse bueno si atiende solo a una fracción de la demanda, aunque consiga que los alumnos tengan altos niveles de aprendizaje. Al menos en los niveles obligatorios, un sistema educativo solo es bueno si todos o casi todos sus destinatarios potenciales son atendidos y consiguen niveles aceptables de aprendizaje.

Para evitar reduccionismos, se explicita la noción de calidad que sustenta este capítulo, basada en el modelo *sistémico*, que resulta adecuado para dar cuenta de la calidad de un *sistema* educativo (INEE, 2006; Martínez Rizo *et al*., 2007), cuyos componentes son:

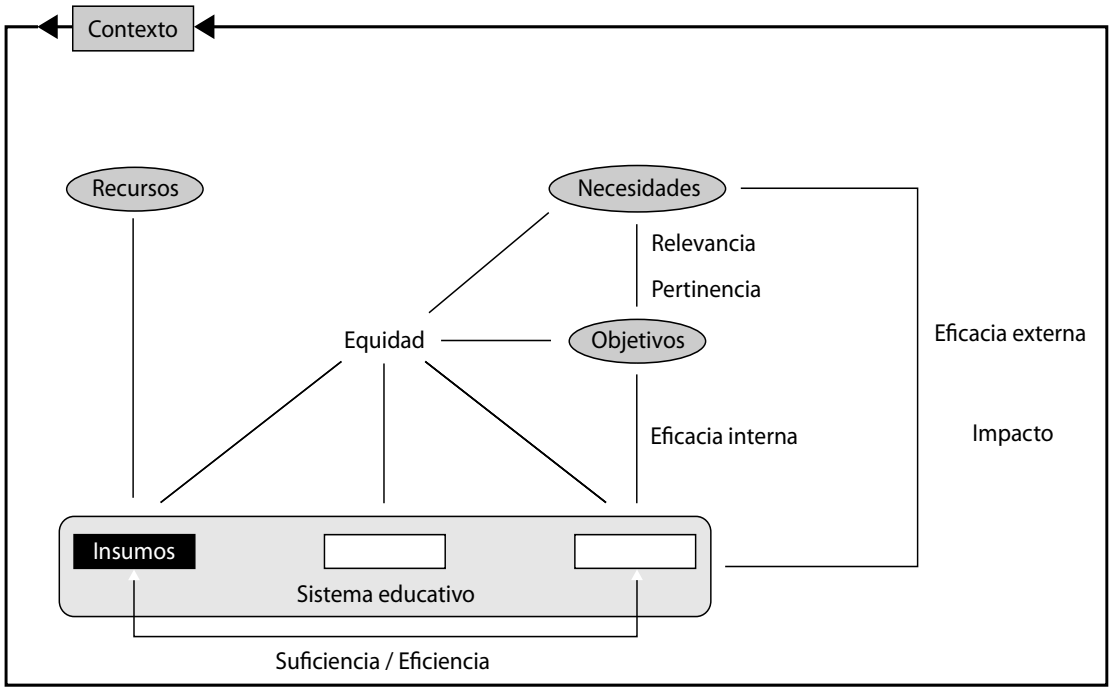
- El *contexto o entorno*, que plantea necesidades al sistema educativo y, al mismo tiempo, le ofrece recursos: la demanda social de educación, su composición y distribución espacial, las condiciones de vida de la población, así como su capital cultural y educativo, son algunos de los elementos del contexto.
- Los *insumos* que recibe el sistema educativo de su entorno permiten su operación e incluyen *recursos* humanos, materiales, tecnológicos y financieros.
- Los *procesos* del sistema se refieren no solo al acceso y trayectoria de los alumnos, sino también a la gestión y funcionamiento de los planteles, incluyendo procesos pedagógicos en el aula; de gestión escolar en cada centro educativo; de gestión o administración en mayor escala en los niveles de zona escolar, de región y nacional.
- Los *productos* o resultados del sistema, de corto o largo plazo. Los primeros incluyen los referidos al logro escolar (conocimientos, habilidades, actitudes y valores), mientras los segundos se relacionan con la influencia de la educación en la transformación cultural y la mejora de los estándares de vida de la población adulta.
- En los sistemas artificiales importan además los *objetivos*, mediadores en la relación entre insumos y productos, orientando la forma en que se organizan los procesos para utilizar los insumos de manera que se generen los productos.

En términos abstractos, la definición de calidad puede expresarse como sigue:

> *La calidad del sistema educativo es la cualidad que resulta de la integración de las dimensiones de pertinencia y relevancia, eficacia interna, eficacia externa, impacto, suficiencia, eficiencia y equidad.*

Esas dimensiones de la calidad se derivan de una relación de congruencia entre los elementos del sistema educativo: pertinencia, relevancia, eficacia externa e impacto se derivan de la relación de objetivos y productos del sistema con las necesidades del entorno; la eficacia interna se desprende de la relación entre productos y objetivos; la eficiencia de la relación entre insumos y resultados; y la equidad de la distribución de recursos y productos educativos. El diagrama siguiente ilustra esta red de relaciones. (Martínez Rizo *et al.*, 2007).

Gráfico. Dimensiones de la calidad de un sistema educativo



La definición abstracta se concreta diciendo que un sistema educativo de calidad:

- Establece un currículo adecuado a las características de los alumnos (*pertinencia*) y a las necesidades de la sociedad (*relevancia*).
- Logra que la más alta proporción posible de destinatarios acceda a la escuela, permanezca hasta el final y aprenda (*eficacia interna y externa*).
- Consigue que los aprendizajes sean asimilados en forma duradera y den lugar a comportamientos sociales fructíferos para la sociedad y los individuos (*impacto*).
- Cuenta con recursos humanos y materiales suficientes (*suficiencia*), y los aprovecha de la mejor manera, evitando despilfarros y derroches (*eficiencia*).
- Considera la desigual situación de alumnos, familias y escuelas, y apoya en especial a quien lo requiera, para que todos alcancen los objetivos educativos (*equidad*).

El concepto de calidad es relativo, porque el juicio depende de los referentes, y dinámico, porque la calidad nunca se alcanza totalmente. La calidad no es un estado, sino una tendencia: la auto-

exigencia permanente y razonable de superación, que no se puede adquirir del exterior, sino que solo surge del interior mismo del sistema.

La noción de evaluación

Un buen sistema de evaluación no deberá prestar atención solo a una u otra dimensión de la calidad; debe atenderlas todas. Algunas veces la evaluación se reduce a aplicar pruebas de rendimiento, lo cual no es adecuado: el nivel de aprendizaje que alcancen los alumnos es una dimensión fundamental, pero no la única, de la calidad.

Es necesario también no aceptar concepciones limitadas de la noción de evaluación. A veces se piensa que se ha *evaluado* el aprendizaje, cuando solamente se ha *medido*. *Evaluar* exige comparar el resultado de la medición con un punto de referencia que establezca lo que debería haber, para llegar a un juicio sobre lo adecuado o inadecuado de lo medido. Por ello se entiende la evaluación como *el juicio de valor que resulta de contrastar el resultado de la medición de una realidad empírica con un parámetro normativo previamente definido*. Por lo mismo, una buena evaluación no se caracteriza solo por las cualidades de una buena medición, como fiabilidad y validez, sino que debe tener esas y otras características (ver INEE, 2006):

- *Carácter comprensivo de la conceptualización que la sustente*, que deberá atender a todas las dimensiones de la calidad.
- *Alto nivel técnico de las mediciones en que se base*, para dar confiabilidad y validez.
- *Pertinencia de los referentes que se definan como parámetros* para contrastar con ellos los resultados de la medición, de modo que las comparaciones tengan sentido.
- *Mesura de los juicios de valor* derivados de contrastar mediciones y parámetros, que evitarán excesos triunfalistas o derrotistas y considerando el contexto.
- *Inclusión de estudios explicativos*, buscando no solo describir la situación, sino también detectar los factores que inciden en ella, para que los actores cuenten con bases sólidas para sustentar decisiones de mejora.
- *Amplitud, oportunidad y transparencia de la difusión de resultados*, que llegue a todos los interesados, en versiones adecuadas a cada uno, para promover su uso.

No sobra señalar que las dimensiones de la calidad no son igualmente susceptibles de medición válida y confiable ni, por lo mismo, de ser evaluadas. Es menos difícil evaluar el aprendizaje que la relevancia del currículo o su pertinencia. Por ello, en muchos casos, solo se podrá aspirar a tener evaluaciones basadas en los juicios más expertos y honestos con que sea posible contar.

Objetos y propósitos de la evaluación

Además de considerar la complejidad de su objeto, y las características de una buena evaluación, los esfuerzos por desarrollar esta deberán tener como punto de partida el principio de que la evaluación no es un fin en sí misma, sino un medio que cobra sentido en función de un propósito más amplio: contribuir a la mejora educativa. Subordinados a ese propósito general hay otros particulares, todos los cuales deberán, a su manera y en su ámbito, contribuir a la mejora de la calidad. Como ejemplos de esa diversidad de propósitos pueden mencionarse:

- *Evaluación de alumnos:* valorar su avance, desde luego, pero también seleccionar a los mejores candidatos para ingresar a cierto nivel; diagnosticar el nivel que tienen al ingreso; asignar becas; dar retroalimentación pedagógica a los docentes; certificar; evaluar el sistema como tal, para la definición de políticas.
- *Evaluación de maestros:* retroalimentarlos sin consecuencias fuertes (formativa); evaluarlos con consecuencias fuertes para su permanencia y promoción (sumativa).
- *Evaluación de escuelas:* evaluar los planteles, para mejora y rendición de cuentas.

La explicitación de posibles propósitos de la evaluación es importante si se tiene en cuenta el principio de que *cada propósito implica un diseño apropiado*, ya que ningún acercamiento da resultados válidos y confiables para todo propósito. Debe reiterarse que el uso inapropiado de cualquier acercamiento puede tener consecuencias negativas, contrarias al propósito básico de que la evaluación favorezca la mejora de la educación (Martínez Rizo, 2007a).

Alcances y limitaciones de acercamientos e instrumentos

Todo instrumento o acercamiento a la evaluación tiene puntos fuertes y débiles, que lo hacen apropiado para algunos propósitos e inapropiado para otros. Por ello, al decidir sobre los instrumentos que se deben usar, es necesario tener en cuenta aspectos como los siguientes:

- El tamaño del universo que se propone evaluar: mientras mayor sea, más difícil de controlar será aplicar instrumentos extensos o de procesamiento laborioso; aplicaciones masivas implicarán instrumentos cortos y que puedan procesarse de forma automática.
- La estructuración de los instrumentos: preguntas de opción múltiple, por ejemplo, no son apropiadas para evaluar aspectos cognitivos complejos o no cognitivos; para ello se necesitan instrumentos que no se limiten a preguntas de opción múltiple.
- La variedad de currículos, que dificulta el uso de pruebas alineadas a ellos; la diferencia entre evaluaciones de impacto alto o bajo, y las implicaciones de evaluar al inicio o al fin del ciclo escolar, uno o varios grados.
- Las exigencias que implican los estudios de valor añadido, que incluyen el uso de diseños que permitan expresar los resultados en una misma escala, el control de las principales variables intervinientes y el seguimiento individual de los alumnos.

Sobre la evaluación de maestros y centros escolares, los resultados de los alumnos en pruebas de rendimiento no pueden tomarse sin más como indicadores confiables de la calidad de un docente o una escuela, por razones precisas:

- Los resultados de los alumnos no dependen solo de la escuela, sino de factores del hogar, la selección de alumnos, el grado en que hay abandono escolar.
- El error de medición produce intervalos de confianza que hacen que la diferencia entre puntajes cercanos no sea significativa; el muestreo tiene consecuencias similares; en censos, el error no derivado del muestreo puede ser mayor.
- Porque en aplicaciones masivas con resultados individuales la extensión de las pruebas es reducida y deben limitarse a algunas áreas y pocos temas curriculares, y el uso de preguntas de opción múltiple hace descartar temas importantes.

Además, asociar consecuencias fuertes a los resultados de las pruebas –como aprobar o no a los alumnos, dar becas o estímulos salariales a docentes, o construir clasificaciones de prestigio de las escuelas– complica la situación. En esos casos hay riesgo de que el trabajo de escuelas y maestros se oriente en función de la prueba, descuidando aspectos importantes del currículo no cubiertos por ella (Martínez Rizo, 2007a).

Debe reiterarse la importancia que tiene el propósito de cada evaluación, ya que de él se derivará el diseño que se ha de utilizar. Un caso que se presenta hoy con preocupante frecuencia y puede tener consecuencias graves para la calidad educativa que se pretende mejorar ilustra esta idea: el que se refiere a las características que deben tener las evaluaciones del aprendizaje de los alumnos que tienen como propósito *dar información sobre ello a cada estudiante*, sus maestros y sus padres, para ajustar en consecuencia sus respectivas acciones para mejorar el aprendizaje ulterior, en contraste con los rasgos que deben tener las evaluaciones del aprendizaje cuyo propósito es *informar sobre la situación promedio del sistema educativo* de que se trate para la toma de decisiones de política. En el primer caso es necesario aplicar las pruebas a todos los alumnos (censo), mientras que en el segundo basta con hacerlo a una muestra representativa, lo que permite usar instrumentos que midan aspectos más complejos.

La complejidad de las evaluaciones integrales

Los límites de los acercamientos en gran escala implican que una evaluación integral, fina, de cualquier objeto, solo pueden hacerla actores cercanos, con la competencia y el tiempo necesarios. Las ventajas de esto, sin embargo, traen consigo inevitablemente la desventaja de la pérdida de perspectiva y la falta de comparabilidad; para atender a estas cualidades son necesarios acercamientos en gran escala, menos finos, pero de cobertura más amplia. Veamos los casos de las evaluaciones de alumnos, maestros y escuelas.

Evaluación de alumnos

Valorar integralmente el desempeño de un solo alumno implica considerar todas las áreas del currículo; los aspectos cognitivos y no cognitivos, simples y complejos; el rendimiento al fin del ciclo, al inicio y a lo largo del mismo; los factores que favorecen u obstaculizan el avance, etc. Una valoración así solo puede ser bien hecha por un profesional que tenga contacto con el alumno a lo largo del ciclo escolar: el maestro. La *evaluación en aula*, sin embargo, aunque es la más completa, tiene una limitación derivada de su misma naturaleza integral y contextualizada: que no puede agregarse en gran escala. Las pruebas estandarizadas permiten comparar grandes grupos de forma confiable; pueden complementar la evaluación del maestro, no sustituirla.

Evaluación de maestros

La valoración integral de la calidad de un maestro solo podrá resultar del contacto amplio de un profesional competente con el evaluando, para reunir pruebas sólidas de su desempeño: ese profesional es el director de escuela. Las evaluaciones de los directores pueden llegar a ser agregables y comparables si se usan metodologías e instrumentos similares y se emplean técnicas de *moderación*. Las pruebas estandarizadas pueden ser útiles para la evaluación de maestros, pero solo sus resultados no son suficientes, por los numerosos factores que inciden en el rendimiento escolar. La aplicación de pruebas a los docentes, si se cuida la validez, aporta elementos valiosos, aunque no capta dimensiones básicas de la práctica docente, que solo pueden apreciarse median-

te acercamientos intensivos. Por ello, la evaluación implica el uso de técnicas como la observación y los portafolios de pruebas.

Evaluación de escuelas

Valorar integralmente la calidad de una escuela solo puede ser el resultado de un amplio contacto con el plantel por parte de otro profesional, capaz de observar, registrar, sistematizar y valorar pruebas de las múltiples facetas de la calidad educativa en ese nivel. Los sistemas educativos suelen contar para ello con la figura del supervisor o inspector. Reducir la evaluación de la calidad de las escuelas a ordenamientos basados en los resultados de pruebas, sin tener en cuenta otros elementos y sin asegurar la confiabilidad y validez de los resultados, es muy inapropiado y llevará a distorsiones perniciosas para el funcionamiento de las escuelas (Martínez Rizo, 2007a).

Los comentarios anteriores se aplican a evaluaciones que servirán para tomar decisiones y emprender acciones que afectarán a sujetos individuales, sean alumnos, maestros o escuelas. Pero si se quiere evaluar en conjunto un sistema de educación, entonces tienen sentido acercamientos a alumnos, maestros o escuelas que utilicen instrumentos que se apliquen a muestras representativas de unos u otras. Esos acercamientos no ofrecen bases sólidas para tomar decisiones graves sobre individuos (como aprobar o reprobar a un alumno, conceder un estímulo a un maestro o una escuela), pero sí elementos valiosos para valorar la calidad del sistema como tal en relación con esos componentes.

La necesidad de un sistema de evaluación

La conclusión de estos someros análisis es que *ningún acercamiento singular puede dar cuenta plenamente de la calidad de un sistema educativo,* pues ninguno podría atender a los diversos propósitos y las circunstancias de sus partes. Debe pensarse en un *conjunto articulado de acercamientos* que formen un *sistema de evaluación* que atienda *como un todo* a las dimensiones de la calidad.

El desarrollo de un conjunto así para un sistema educativo implicará una gama de acercamientos evaluativos que cubran distintos objetos (alumnos, docentes, directivos, planteles, etc.); atiendan a varios propósitos (diagnóstico, retroalimentación, rendición de cuentas, selección, certificación...), e incluyan tanto acercamientos de gran escala como en pequeña escala. Solo combinando varios acercamientos cuyos alcances y limitaciones se complementen podrá tenerse una visión suficientemente completa de ese todo tan complejo que es la calidad de un sistema educativo, de manera tal que sea posible formular juicios de valor que sustenten de manera razonable las decisiones de las autoridades en los diferentes niveles.

PROPUESTA DE MODELO PARA UN SISTEMA DE EVALUACIÓN

Para atender a los diferentes aspectos de la calidad de un sistema educativo, un sistema de evaluación no deberá limitarse a la aplicación de pruebas estandarizadas. Además de tales pruebas –si son de buena calidad y se usan de manera apropiada– se deberán atender otros aspectos del sistema educativo, sus alumnos, maestros y escuelas. Para evaluar al sistema educativo como tal, se necesitan indicadores de su cobertura y su eficiencia terminal y del contexto socioeconómico en que se inscribe, a partir de las estadísticas educativas y sociales. Un sistema de evaluación deberá comprender un subsistema de evaluación de alumnos, uno de evaluación de maestros, otro de evaluación de escuelas y uno de indicadores educativos (Martínez Rizo, 2007a).

El subsistema de evaluación de alumnos

Sin olvidar el carácter insustituible de la evaluación que hacen los maestros en el aula, las pruebas en gran escala pueden ser un complemento interesante, si se hacen bien y se tiene conciencia de sus alcances y límites; permiten elaborar juicios sobre la relación que hay entre lo que se pretende enseñar (*currículo prescrito*) y lo que los estudiantes aprenden (*currículo alcanzado*), con la mediación del *currículo implementado*, que a veces se denomina *oportunidad de aprendizaje*. Para cumplir con el propósito de diagnóstico de la situación de un sistema educativo en cuanto al aprendizaje de los alumnos, para retroalimentar las políticas al respecto, unas buenas pruebas en gran escala deberán caracterizarse por rasgos como los siguientes (Martínez Rizo, 2007b):

- *Tener un referente bien definido*: el currículo, un perfil del egresado, un conjunto de competencias básicas, etc. Si están alineadas al currículo, las evaluaciones darán insumos valiosos para orientar la formulación de políticas curriculares, así como de formación y actualización de profesores.

- *Estar construidas con referencia a un criterio o dominio*: diseñadas para conocer el desempeño de los alumnos en relación con contenidos bien definidos, lo que deberá sustentar la formulación de juicios de tipo absoluto, en términos de adecuado-inadecuado, suficiente-insuficiente o similares, y no solo juicios relativos, respecto al nivel promedio obtenido por el conjunto de los estudiantes evaluados.

- *Tener un diseño matricial y aplicarse en ciclos.* La amplitud de los contenidos que se van a evaluar hace necesario dividir las pruebas en partes, para que cada alumno responda solo algunas, con un diseño que arroje resultados sobre el contenido completo de las pruebas a partir de los resultados del conjunto de los alumnos. Con pruebas así, la variedad de áreas curriculares podrá evaluarse con un ciclo multianual bien diseñado, pero no podrán cubrir bien aspectos cuya importancia no debe subestimarse.

- *Comprender preguntas de diverso tipo.* Las más frecuentemente usadas –las de *opción múltiple*– tienen la ventaja de que pueden procesarse de manera mecánica, pero difícilmente permiten evaluar competencias complejas, que requieren del uso de preguntas de respuesta construida o de ejecución, que implican procesos de calificación a cargo de personas capacitadas, lo que excluye aplicaciones masivas.

- *Aplicarse a muestras de escuelas y alumnos.* Los rasgos anteriores implican que la aplicación de estas pruebas no podrá ser de tipo censal, sino que deberá hacerse a muestras de planteles y alumnos, definidas de tal manera que sean representativas del conjunto del sistema nacional de educación y de sus principales subsistemas.

- *Acompañarse de cuestionarios de contexto.* Para explicar los diferenciales de logro, las pruebas han de acompañarse de instrumentos para recabar información sobre las características personales, familiares y escolares de los estudiantes; las de los docentes y de las aulas en las que laboran, y las del director y las escuelas mismas.

Si se quiere tener resultados comparables de cada centro escolar, e incluso de cada alumno, lo que puede ser útil para efectos de retroalimentación para alumnos, padres de familia, maestros y escuelas, serán necesarias aplicaciones censales, que tendrán que cubrir un espectro menor de contenidos, que se puedan evaluar mediante preguntas cerradas.

El desconocimiento de los alcances y límites de las pruebas en gran escala lleva a veces a usos inadecuados. Se aplica un número excesivo de pruebas, en demasiados grados y con demasiada frecuencia, sin advertir que a medida que aumenta el tamaño de una aplicación crece también el riesgo de errores no derivados del muestreo y resultados no confiables. Tampoco se tienen en cuenta las exigencias que implica desarrollar y aplicar pruebas de buena calidad; se plantean a los responsables de ello tareas considerables en tiempos demasiado cortos para que se hagan bien (ver Ravela *et al*., 2008).

Aplicaciones censales pueden ser útiles para retroalimentar a escuelas y alumnos de forma individual, pero es suficiente hacerlas en algunos grados y áreas y con intervalos de tiempo amplios. Aplicar pruebas a muchos grados cada año no es necesario y trae consigo problemas de calidad, además de costos elevados y de una malsana presión sobre las escuelas, sin olvidar los usos inapropiados que suelen darse a los resultados.

El subsistema de evaluación de maestros

Evaluar bien a un docente implica observar su trabajo en el aula y revisar cómo prepara sus clases y retroalimenta a sus alumnos. Por ello, las mejores formas de evaluar el desempeño docente son variantes de la técnica de *portafolio de pruebas*, que supone una importante inversión de tiempo y, por lo mismo, hacen difícil su uso en gran escala. Se han mencionado ya las precauciones que deberán tenerse para utilizar los resultados de los alumnos de un docente para evaluar a este, lo que no se opone a una utilización cuidadosa de tal información, junto con otras.

Otros acercamientos pueden aportar elementos útiles, como evaluaciones estandarizadas de los conocimientos de los docentes sobre las áreas a su cargo y encuestas que recojan la opinión de los alumnos sobre sus maestros. Pese a la resistencia que suelen enfrentar, esas encuestas pueden dar información valiosa. La resistencia es explicable en el caso de evaluaciones de las que se derivan decisiones sobre individuos, que no se sustentan bien solo con instrumentos de aplicación en gran escala, que pueden dar información valiosa para la valoración general del sistema educativo, con aplicaciones muestrales.

Los alumnos pueden informar confiablemente sobre la cobertura del programa por parte del docente, la atención que presta a su grupo e, incluso, una dimensión tan importante del desempeño como es la regularidad de la asistencia del maestro. En los sistemas educativos consolidados se da por hecho que todo grupo tiene siempre un docente que lo atienda; en sistemas menos consolidados las ausencias de los maestros pueden ser frecuentes, por diversas razones, como el aislamiento del centro escolar, la inexistencia de maestros que puedan suplir a los ausentes, la rotación de personal e incluso fallas administrativas y vicios burocráticos y de carácter gremial.

La noción de *oportunidad de aprendizaje*, que en un sistema consolidado alude a variantes menores en el uso del tiempo por el docente, y sus estrategias didácticas, en sistemas menos consolidados comienza con la regularidad de la asistencia del maestro, prerrequisito esencial de la calidad sobre el que no suele haber información. Una buena evaluación de maestros en el nivel del sistema integrará información sobre esos aspectos con datos agregados de evaluaciones del desempeño docente hechas por instancias de dirección escolar y supervisión, con mediciones de conocimientos profesionales y con indicadores estadísticos sobre la composición y el perfil de la planta docente.

El subsistema de evaluación de escuelas

El trabajo de alumnos y maestros no ocurre de forma aislada, tiene lugar en el marco de una escuela, que es un ambiente colectivo con peso específico, por características organizacionales que los estudios sobre escuelas eficaces han identificado. En contra de lo que se pensó en los años sesenta y setenta del siglo pasado, tras la difusión del estudio dirigido por James Coleman, que destacaba la influencia de factores del entorno social sobre los resultados escolares, el peso de la escuela en esos resultados no es insignificante.

La escuela sí cuenta y puede hacer una diferencia. Además de la evaluación de alumnos y maestros, la de la escuela como unidad organizacional es importante, tal vez incluso la más importante desde el punto de vista de las políticas educativas, ya que la mejora de la calidad del sistema pasa por la mejora en cada aula de cada plantel; por ello la transformación de la dinámica de funcionamiento de cada escuela es fundamental.

Una vez más, una evaluación fina y razonablemente completa de una escuela solo puede resultar del trabajo de actores que tengan un contacto cercano con ella, normalmente las instancias de supervisión o inspección, más autoevaluaciones a cargo de los docentes, con el director a la cabeza en relación con el plan de desarrollo o mejora de cada plantel.

Como se ha advertido, las evaluaciones en gran escala del aprendizaje de los alumnos no pueden utilizarse como indicadores directos de la calidad de cada escuela, si no se han utilizado diseños apropiados para ello, que implican estudios del valor agregado por la escuela (ver Martínez Rizo, 2006b). Aun sin ello, los resultados de evaluaciones de aprendizaje pueden ser valiosos como indicadores parciales, si se manejan con cuidado, junto con los elementos de contextualización de que se disponga. Convendrá también hacer estudios particulares sobre ciertos recursos y procesos de las escuelas.

El subsistema de indicadores

Ninguna evaluación particular puede dar cuenta por sí sola de ese elusivo objeto que denota la expresión *calidad del sistema educativo*; hay que combinar acercamientos como los anteriores, en relación con alumnos, maestros y centros escolares. Otros aspectos de la calidad se pueden captar con indicadores derivados de las estadísticas tradicionales, como cobertura, eficiencia terminal y deserción escolar. Esos indicadores se deberán complementar con los que se deriven de las evaluaciones de alumnos, maestros y escuelas. En este sentido, el *subsistema de indicadores educativos* es el elemento integrador del *sistema de evaluación*.

Los listados de indicadores de algunos sistemas educativos parecen simples concentrados de datos producidos a partir de la información estadística con la que se cuenta en un momento dado, independientemente de su relevancia y sin identificar las lagunas que tenga. Esos listados no son verdaderos sistemas, puesto que carecen de una estructura basada en alguna idea clara de lo que se busca con ellos. La construcción de un verdadero sistema de indicadores debe hacerse a partir de una idea clara de lo que se quiere obtener, que dé unidad al conjunto. Esa idea rectora debe ser *una noción precisa de la calidad del sistema educativo*, que es lo que se quiere evaluar.

A partir de esta noción podrán precisarse indicadores particulares, y luego los datos que se requieren para construirlos. El que muchos de esos datos no estén disponibles no deberá impedir

que el sistema los considere, lo que hará que se emprendan las acciones que se requieran para llegar a contar con la información necesaria en algún momento.

Este enfoque *sistémico* del desarrollo de indicadores educativos producirá un número considerable de ellos, lo que no es un defecto, sino una consecuencia de la complejidad del objeto de análisis. Es posible, sin embargo, adoptar otro enfoque, que se puede llamar *estratégico*. En esta perspectiva se puede pensar en un número reducido de indicadores que permitan monitorear el avance en dirección a metas precisas, que se han definido como importantes. Es el caso de los indicadores desarrollados para dar seguimiento a las metas de la Unión Europea. El sistema INES (Indicators of National Education Systems) de la OCDE es un buen ejemplo del primer enfoque.

A diferencia de lo que ocurre para la construcción de pruebas estandarizadas, no hay metodologías ampliamente aceptadas para formular y construir indicadores. Por ello se presentan en seguida los principales elementos de la metodología desarrollada por el Instituto Nacional para la Evaluación de la Educación de México, a partir de la experiencia del propio INEE, y recogiendo elementos de la experiencia internacional (Martínez Rizo *et al*. 2007; Shavelson *et al*., 1989; Comisión Europea, 1999).

Según dicha metodología, la construcción de un sistema de indicadores debe comenzar precisando el marco conceptual que organizará las partes en un verdadero sistema, en vez de una serie de indicadores aislados. Una vez establecida la estructura del sistema se elabora un primer listado de indicadores que cubra todas las dimensiones consideradas, aunque para algunos de ellos no se cuente por el momento con información.

La segunda etapa del proceso de diseño se refiere al desarrollo de cada indicador, lo que significa precisar su nombre; su definición y su ubicación en el sistema; las fuentes de obtención de la información; la fórmula de cálculo y, en su caso, los criterios para su estimación; las desagregaciones posibles (por ejemplo, por género, por región, por modalidad de escuela, etc.), y los elementos necesarios para la interpretación de los resultados.

Tras lo anterior –y antes de proceder a organizar costosos procesos de recolección de información para alimentar un sistema de indicadores–, cada uno de los elementos que lo forman debe someterse a escrutinio, con un procedimiento de jueceo entre expertos y usuarios, para valorar *a priori* el grado en que cada indicador parece satisfacer los criterios de calidad aplicables, que son, además de los de validez y confiabilidad, otros de carácter técnico (comparabilidad, estabilidad temporal, actualidad, sensibilidad) y algunos más de tipo práctico: factibilidad de obtención de la información necesaria, importancia y utilidad para la toma de decisiones, y claridad de su interpretación.

El análisis inicial *a priori* del conjunto de indicadores diseñado deberá involucrar a un número razonable de jueces, de tres tipos: expertos o especialistas de formación y experiencia teórica o disciplinaria sobre los temas cubiertos por el sistema; usuarios especializados, en el sentido de responsables del sistema educativo con experiencia de nivel alto y medio, y usuarios de base, como supervisores, directores y maestros. Tras lo anterior podrá procederse a una primera aplicación del sistema, alimentándolo con la información necesaria, previamente existente o generada *ex profeso*, y con los procesos de depuración y cálculo necesarios para obtener los primeros valores de cada indicador.

Tras la primera aplicación se podrá hacer un nuevo análisis de la calidad de cada indicador y del sistema en conjunto, *a posteriori*, revisando la consistencia de los datos y contrastándolos con otros comparables, por ejemplo, de investigación independiente. Se podrá valorar también en qué medida la información es interpretada correctamente y utilizada para sustentar acciones de mejora.

CONCLUSIÓN

La difusión de pruebas de rendimiento en gran escala que hoy se observa en muchos sistemas educativos se entiende, en parte, por la creciente insatisfacción de personas interesadas por la educación respecto a los niveles de aprendizaje de los alumnos. No pocos de los que experimentan tal insatisfacción, sin embargo, tienen una visión simplista tanto de las causas de la situación como de la forma de valorarla y corregirla.

En cuanto a causas, son frecuentes los diagnósticos que no toman en cuenta los muchos factores de las escuelas y su contexto que influyen en la calidad. En cuanto a los medios de valorar la calidad, las visiones simplistas los reducen a la aplicación de pruebas de rendimiento, sin distinguir su enfoque y nivel técnico. Y en cuanto a las estrategias para remediar la situación, se proponen medidas simples de las que se espera demasiado, como asignar recursos a las escuelas según los resultados de sus alumnos en las pruebas, los bonos educativos (*vouchers*), la privatización de las escuelas públicas, etc.

En este contexto se observa una proliferación de pruebas con las que se pretende valorar la calidad de las escuelas, aunque los instrumentos no hayan sido diseñados de manera apropiada para tal propósito y aunque tengan limitaciones técnicas que ponen en tela de juicio su validez y su confiabilidad. A partir de ello se producen dos tipos de reacciones: una, los maestros se ven impulsados a orientar su trabajo pedagógico en función de los contenidos de las pruebas, ya que serán el medio privilegiado para valorar su trabajo; y otra, entre los mismos maestros y entre los estudiosos de la educación, se genera un rechazo absoluto de las pruebas, por considerarlas intrínsecamente inadecuadas.

La postura que se sostiene en este trabajo es diferente. Se acepta que, en muchos casos, la calidad de la educación no es buena, en especial en relación con los países de mejores resultados; antes de calificarla como catastrófica, sin embargo, se considera también que hay que tener en cuenta que las causas son complejas y suficientemente conocidas como para que los resultados no resulten sorpresivos para quien tenga una visión amplia del contexto nacional e internacional. Se considera que buenas evaluaciones son necesarias para apoyar esfuerzos de mejora, pero que debe tenerse mucho cuidado para no caer en el error que hoy acecha a los sistemas educativos: el desarrollo de evaluaciones no adecuadas. Como toda herramienta, la evaluación puede ser utilizada bien o mal. Un uso inadecuado de un instrumento tan poderoso puede tener consecuencias negativas graves, que las personas preocupadas por la educación deberíamos tratar de evitar.

Las evaluaciones internacionales

Enrique Roca

En el presente capítulo se efectúa un repaso de los estudios internacionales de evaluación en los que han estado implicados buena parte de los países europeos y americanos. Se trata de revisar la evolución de estos estudios desde su nacimiento a comienzos de la década de los sesenta, las razones que los impulsaron, el objeto de estudio, las poblaciones a las que se han dirigido, los instrumentos utilizados, las características principales de los análisis realizados, la presentación de sus resultados y la utilidad de los mismos para el conocimiento de las realidades educativas. El lector encontrará una reflexión más detenida sobre las evaluaciones y la definición de objetivos y de políticas educativas en el capítulo de Alejandro Tiana en este libro.

Además de la valoración que realizan los maestros de los aprendizajes de sus alumnos, desde la primera mitad del siglo xx la evaluación de los sistemas educativos se convirtió en una preocupación fundamental de los responsables de la educación. La elaboración de indicadores sociales sobre el estado de la educación de la población (analfabetismo) o las primeras estadísticas sobre el número de escuelas y maestros o los porcentajes de alumnos escolarizados y egresados son los primeros pasos dados para conocer el estado de los sistemas educativos. Pero la evaluación a partir de los rendimientos alcanzados por los alumnos cobró un importante impulso desde la segunda mitad del siglo xx, sobre todo en Estados Unidos, una vez que se pusieron a punto los instrumentos de medida adecuados, como señala Martinez Rizo en su ensayo en este libro.

LOS PRIMEROS ESTUDIOS INTERNACIONALES: LOS TRABAJOS DE LA IEA Y DEL LLECE

Los primeros pasos en la preparación y realización de estudios internacionales comparados de evaluación de los sistemas educativos se dan en los años cincuenta; en 1958 se inician en el Instituto de la UNESCO para la Educación en Hamburgo los primeros contactos que dan lugar, en 1967, al nacimiento formal de la IEA (International Association for the Evaluation of Educational Achievement). A los estudios propuestos por esta asociación se fueron incorporando institutos de investigación y evaluación educativa que incluían universidades y organismos dependientes de Ministerios de Educación hasta alcanzar la cifra actual de 62 instituciones y países implicados (IEA, 2007), entre los que se encuentran, además de España y Portugal, Argentina, Brasil, Chile y México.

El primer trabajo de la IEA se realizó entre 1959 y 1962 con muestras de estudiantes de 13 años de 12 países y abarcó cinco áreas: matemáticas, comprensión lectora, geografía, ciencias y una habilidad no verbal. Independientemente de los resultados académicos y prácticos, el estudio mostró que era posible realizar mediciones internacionales a gran escala y que distintos centros de investigación de diferentes países podían construir de común acuerdo cuestionarios y conjuntos de pruebas para realizar análisis de resultados de los alumnos, independientemente de que estos trabajaran en distintas lenguas y con diferentes currículos y contextos educativos.

El éxito de este estudio preliminar animó a iniciar en 1964 el primer estudio internacional sobre matemáticas (FIMS), en el que participaron doce países, entre los que no se encontraba ninguno latinoamericano. Se recogieron datos sobre dos poblaciones de estudiantes de 13 años y alumnos de estudios preuniversitarios. En 1970-1971 la IEA realizó un nuevo grupo de estudios en el que se examinaron seis materias: comprensión lectora, ciencias, literatura, francés e inglés como lenguas extranjeras y educación cívica, con estudiantes de 10 y de 14 años (edad que por entonces coincidía con el final de la educación obligatoria de la mayoría de los países participantes). Este estudio incluyó aspectos relacionados con el interés, la motivación y las actitudes de los alumnos, los métodos de enseñanza y las prácticas escolares, y se completó en 1985 con un estudio sobre expresión escrita.

A partir de los años ochenta cobraron mayor intensidad estos estudios internacionales de evaluación patrocinados por la IEA (2007a), sobre las áreas señaladas y sobre otras nuevas que se fueron incorporando sucesivamente: SIMS, SISS, TIMSS (a partir de 1999 la T significa *trends*), PIRLS, ICCS... Estos estudios se realizaron con alumnos escolarizados en distintos grados de enseñanza y de distintas edades, incluida la educación infantil, la educación primaria y la educación secundaria.

Estudios de evaluación de la IEA	
1959-1962	*Pilot Twelve-Country Study.*
1964	*Primer estudio internacional sobre matemáticas (First International Mathematics Study-FIMS).*
1970-1971	*Estudio sobre comprensión lectora, ciencias, literatura, francés e inglés como lenguas extranjeras y educación cívica.*
1985	*Estudio sobre Written Composition.*
1980-1981	*SIMS: Second International Mathematics Study.*
1983-1984	*SISS: Second International Sciencia Study.*
1987-1993	*COMPED: The Computers in Education Study.*
1991	*The Use of Computers in Education Worldwide.*
1993	*The IEA Study of Computers in Education.*
1995	*The IEA Reading Literacy Study. Technical Report.*
1995	*TIMSS: Third International Mathematics and Science Study.*
1999, 2002	*SITES: The Second Information Technology in Education Study.*
1994-2002	*CIVED: Civic Education Study.*
1999, 2003, 2007	*TIMSS: Trends in Mathematics and Science Study.*
2001, 2006	*PIRLS: Progress in Reading Literacy Study.*
	ICCS: Estudio Internacional sobre Educación Cívica y Ciudadana.
	TEDS-M: Estudio Internacional sobre la Formación Inicial del Profesorado de Matemáticas.

A partir de los años noventa, a los trabajos internacionales de evaluación de la IEA se han sumado los realizados por distintas organizaciones internacionales: UNESCO, OCDE y Unión Europea, entre otras. Se han ampliado los objetos de estudio: además del rendimiento de los alumnos en

distintas áreas del currículo, y en distintas edades de la educación primaria y secundaria, se han propuesto estudios sobre distintas competencias básicas (OCDE, UE), actitudes y valores de los alumnos (IEA, UE), formación del profesorado (IEA), y procesos de enseñanza y aprendizaje, basados en cuestionarios a los profesores (OCDE), o sobre otras etapas de la formación a lo largo de la vida. En los cuadros que acompañan a este texto se puede observar una relación de los principales estudios de las instituciones señaladas. A los recientes estudios de UNESCO, OCDE e IEA se han ido incorporando distintos países latinoamericanos, como se señala a continuación en cada caso.

En 1994 se fundó el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), coordinado por la Oficina Regional para América Latina y el Caribe (OREALC) de UNESCO. Entre 1995 y 1997, el LLECE realizó el Primer Estudio Regional Comparativo y Explicativo (PERCE), y entre 2002 y 2008 ha realizado el Segundo Estudio Regional Comparativo y Explicativo (SERCE), de los cuales el lector encontrará detallada información en el capítulo de Guillermo Ferrer en este libro (LLECE, 2008).

Los estudios de evaluación en América Latina del LLECE, UNESCO-OREALC	
1997	*PERCE: Primer Estudio Regional Comparativo y Explicativo.*
2007	*SERCE: Segundo Estudio Regional Comparativo y Explicativo. 3.º y 6.º nivel en lenguaje, matemáticas y ciencias.*

Finalmente, debe señalarse otro proyecto importante en el plano internacional, el SACMEQ (*Southern and Eastern Africa Consortium for Monitoring Educational Quality*), en el que están implicados 15 países del este y sur de África y el Instituto Internacional de Planeamiento de la Educación de la UNESCO (IIEP). El consorcio fue creado formalmente en 1995 y realizó un primer estudio en 1998, el SACMEQ I, y uno segundo, el SACMEQ II en 2000-2003, en el que participaron 14 países.

LA EVALUACIÓN DE COMPETENCIAS BÁSICAS: PISA Y LOS ESTUDIOS DE LA UNIÓN EUROPEA

La OCDE inició el proyecto PISA (Programa para la Evaluación Internacional de Alumnos) en 1997 con el propósito de ofrecer resultados sobre rendimiento educativo de los alumnos de 15 años a partir de la evaluación, cada tres años, de su grado de adquisición de tres competencias básicas: la comprensión lectora (MEC. Instituto de Evaluación, 2007a), la matemática y la científica (OCDE, 2005). Se trataba de completar con estos resultados educativos el panorama de indicadores educativos que venía publicando la OCDE desde el inicio del Proyecto INES en 1992. Pero, sobre todo, PISA estudia la evolución de los sistemas educativos a través de los logros de los alumnos. Las características fundamentales de PISA han sido su orientación política y su innovador concepto de *competencia básica*, que tiene que ver con la capacidad de los estudiantes para extrapolar lo aprendido y aplicar sus conocimientos ante nuevas circunstancias, su relevancia para el aprendizaje a lo largo de la vida y su regularidad.

Además PISA, como buena parte de los estudios internacionales recientes, realiza una evaluación de los *factores* que pueden estar asociados con los distintos niveles de competencia, que se abordan más adelante. En la edición de 2000 participaron 43 países (entre ellos Argentina, Brasil,

Chile, México y Perú) y en la del 2003, 41 países (entre los cuales se encontraban Brasil, México y Uruguay). En la edición de 2006 han participado 57 países, incluidos los 30 de la OCDE y otros 27 países asociados, y entre ellos los siguientes países latinoamericanos: Argentina, Brasil, Chile, Colombia, México y Uruguay (OCDE, 2008). Estos países, junto con España y Portugal, han constituido el Grupo Iberoamericano de PISA (GIP), que viene analizando conjuntamente los trabajos de PISA desde la perspectiva iberoamericana y está elaborando un informe regional con los resultados PISA 2006.

La OCDE ha puesto también en marcha dos estudios de evaluación, TALIS (Estudio Internacional de Enseñanza y Aprendizaje), donde participan, además de España y Portugal, Brasil y México, y PIACC (Programa para la Evaluación Internacional de las Competencias de los Adultos), que comportan un trabajo de campo dirigido a profesores y a adultos. Además, la OCDE ha propuesto otros estudios de evaluación que se encuentran en una fase de informes nacionales, análisis y propuestas de futuro: liderazgo (Pont, 2008), reconocimiento de los aprendizajes no formales e informales, o el proyecto sobre el desarrollo de modelos de valor añadido en los sistemas educativos.

Estudios de evaluación de la OCDE	
2000, 2003, 2006	*PISA: Programa para la Evaluación Internacional de Alumnos.*
2005	*ALL: Adult Literacy and Life Skills Survey.*
	TALIS: Estudio Internacional de Enseñanza y Aprendizaje.
	PIACC: Programa para la Evaluación Internacional de las Competencias de los Adultos.
	RNFIL: Reconocimiento de los Aprendizajes No Formales e Informales.
	Liderazgo en los centros educativos.
	Desarrollo de modelos de valor añadido en los sistemas educativos.

Desde la presentación del informe Delors (1996), en la Unión Europea se puso en marcha la definición de las competencias clave. La *Estrategia de Lisboa* se concretó, en educación, en la definición de los objetivos educativos para 2010 y en la recomendación a los países de que se trabaje para que los alumnos puedan adquirir en la educación obligatoria las competencias clave. Además, la Unión Europea ha puesto en marcha iniciativas para evaluar los rendimientos obtenidos tomando como referencia, precisamente, la adquisición de las competencias por parte de los alumnos (UE, 2006). Esta propuesta ha venido acompañada de un intenso programa de evaluación educativa que consiste en dos procesos convergentes: el estímulo y el apoyo a los países miembros para su participación en estudios de la OCDE y de la IEA, y la realización de estudios propios.

Entre los estudios de la Unión Europea se pueden señalar el relativo a la evaluación de la competencia *aprender a aprender*, con el que además se pretende contribuir a la construcción de un nuevo indicador educativo europeo; la elaboración del *Banco Europeo de Ítems de Anclaje para la Evaluación de Destrezas en Lenguas Extranjeras*, basado en los niveles de rendimiento establecidos en el *Marco Común Europeo de Referencia para las Lenguas,* o el proyecto sobre un *Indicador Europeo de Competencia Lingüística*, que debe permitir conocer y estimular el aprendizaje de idiomas, en particular el aprendizaje de dos lenguas distintas a la materna desde una edad muy temprana (UE, 2006).

Estudios de evaluación de la Unión Europea
EBFALS: Banco Europeo de Ítems de Anclaje para la Evaluación de Destrezas en Lenguas Extranjeras.
Indicador Europeo de Competencia Lingüística.
Aprender a aprender.

LAS EVALUACIONES INTERNACIONALES DE LOS SISTEMAS EDUCATIVOS A PARTIR DEL RENDIMIENTO DE LOS ALUMNOS

En este apartado se resumen algunos de los aspectos y características esenciales de los estudios internacionales de evaluación, centrados en el rendimiento de los alumnos y los principales factores asociados al mismo, mencionados en el apartado anterior. Conviene resaltar de entrada que la evolución de estos estudios ha sido notable, como lo ha sido la mejora de los instrumentos de medida, que han alcanzado elevado rigor y fiabilidad (Roca, 2008).

Poblaciones y muestras

En primer lugar, hay que resaltar que los estudios internacionales se han dirigido a diferentes poblaciones de alumnos. En unos casos, se ha centrado el estudio en determinados grados de primaria: PIRLS, Progreso en Capacidad de Lectura en 4.º (MEC, Instituto de Evaluación, 2007b), y LLECE en 3.º, 4.º y 6.º; o en secundaria: TIMSS, Estudio Internacional de Tendencias en Matemáticas y Ciencias, en 8.º, ICCS (IEA, 2007b), Estudio Internacional sobre Educación Cívica y para la Ciudadanía, en 8.º, independientemente de la edad de los alumnos. En las comparaciones internacionales, algunos expertos plantean el inconveniente que supone, al evaluar un grado determinado, valorar a alumnos con distinta edad y número de años de aprendizaje acumulados y las diferencias que este hecho puede introducir, que no podrán ser atribuidas al rendimiento de cada uno de los sistemas educativos comparados.

En otros casos, el estudio se ha realizado con alumnos de una determinada edad (PISA a los 15 años), o al finalizar una etapa educativa (EBFALS, el Banco Europeo de Ítems de Anclaje para la Evaluación de Destrezas en Lenguas Extranjeras, al acabar la secundaria obligatoria). Esta decisión permite comprobar la eficacia de los distintos sistemas a la hora de garantizar la adquisición de las competencias básicas por parte de sus alumnos en un momento próximo al final de la educación obligatoria en la mayoría de los países. Este tipo de análisis pone de manifiesto que los rendimientos de los alumnos que siguen determinados itinerarios formativos o que padecen la repetición de curso son inferiores que los de otros compañeros que a la edad considerada han seguido con éxito los programas generales de estudio que habitualmente dan acceso a estudios académicos postobligatorios.

En todos los casos, las muestras de alumnos seleccionados para representar a cada país (en su caso región, estado federado o comunidad autónoma) son tales que deben ofrecer una confianza no inferior a la establecida como garantía en cada estudio. El tamaño de las muestras está condicionado igualmente por el diseño matricial que suele adoptarse para la aplicación de diferentes cuadernillos de las pruebas. El muestreo se realiza habitualmente mediante una primera selección aleatoria de los centros educativos de cada país participante (o región). En una segunda etapa, en cada uno de los centros seleccionados se eligen aleatoriamente los alumnos que corresponda.

Tipos de ejercicios

Los ejercicios que realizan los alumnos suelen consistir en una prueba de lápiz y papel con cuadernos distintos, con el contenido parcialmente solapado según la técnica de muestreo matricial, para que se cubra un campo más amplio de contenido, pero sin superar el tiempo previsto de prueba individual. De este modo, cada alumno responde a un conjunto de ítems tal que el ejercicio no sea para él excesivamente largo, pero que, al mismo tiempo, la prueba en su conjunto pueda contar con un número suficiente de respuestas para cada ítem y un número de ítems también suficiente para asegurar que se mide realmente lo que se pretende y se hace de modo objetivo.

Por otra parte, los alumnos suelen responder a un cuestionario de datos personales y familiares, que incluye sus opiniones y actitudes ante el estudio y ante la atención que reciben en el centro escolar y en el aula. Estos cuestionarios son esenciales para obtener información complementaria sobre distintos factores asociados a los aprendizajes, como son los contextos sociales, económicos y culturales de alumnos y centros, el clima del aula, los procesos educativos, etc.

En muchas ocasiones, las pruebas que realizan los alumnos son complementadas con distintos cuestionarios para el director del centro o miembros del equipo directivo, para los profesores de los alumnos participantes en la prueba y para sus familias, cuestionarios todos que permiten ampliar y matizar la información relativa a los factores asociados proporcionada por los alumnos.

Las *preguntas* de la prueba de rendimiento pueden ser de distinto tipo. Por un lado, hay preguntas cerradas en dos versiones: simples, o de elección múltiple, frecuentemente entre cuatro opciones distintas de respuesta, y complejas, que combinan diferentes opciones de respuesta binaria (sí/no; verdadero/falso). También hay preguntas abiertas, de respuesta elaborada, corta o larga, que necesitan la intervención de correctores bajo pormenorizados criterios de valoración. Por último, las preguntas sobre actitudes suelen utilizar escalas en las que los alumnos identifican el valor más acorde con sus disposiciones, actitudes o valoración personales (*escalas Likert*).

En la mayoría de los estudios, las preguntas no se presentan aisladas, sino formando grupos o unidades de evaluación bajo una presentación textual o gráfica común, (estímulo), que propone al alumno una situación cotidiana como la que se puede encontrar en la vida real, sobre la que se le formulan distintas cuestiones. Esta presentación permite minimizar el efecto de los cambios de contexto de las preguntas y también posibilita que el estímulo común pueda ser mejor explotado.

Áreas curriculares y competencias básicas

Desde el inicio, los estudios internacionales de evaluación trataron de medir el aprendizaje de los alumnos mediante los resultados que obtienen o, dicho de otro modo, mediante el análisis de sus conocimientos y destrezas en determinadas áreas del conocimiento.

En esos primeros estudios se consideraron la comprensión lectora en la lengua materna, las matemáticas y, muy frecuentemente, las ciencias naturales, como las áreas principales sobre las que organizar las pruebas que deben responder los alumnos. Las dos primeras áreas se caracterizan por ser instrumentales y universales y ocupar un lugar destacado en los currículos de todos los países. Las ciencias no disfrutan exactamente de las mismas características que lectura y matemáticas, pero sí suelen presentar un tratamiento bastante común en los currículos y una pretensión de objetividad elevada, circunstancias imprescindibles para una medida más precisa y objetiva. Estas tres áreas se prestan bien a su evaluación con instrumentos estandarizados. Otras áreas han

sido objeto de estudio, como la geografía o los idiomas extranjeros, pero con menor frecuencia, debido sobre todo al diferente tratamiento que reciben en los currículos de los distintos países, a sus peculiaridades regionales o nacionales o a la importancia que, con mayor o menor justificación, se otorga a los distintos saberes.

A partir de los años noventa fue cobrando fuerza en la evaluación la tendencia a considerar, además de los aspectos meramente cognitivos de las distintas áreas del currículo, las denominadas competencias básicas. Al evaluar competencias básicas se trata de medir no tanto los contenidos que los alumnos han aprendido, como su capacidad para extrapolar sus aprendizajes y aplicar sus conocimientos y habilidades a nuevas situaciones y contextos, es decir, aplicar lo que aprenden en la escuela y fuera de ella y tomar decisiones fundadas. Este es el enfoque de PISA de la OCDE (OCDE, 2005).

Por otra parte, la propuesta de la Unión Europea de la incorporación de las competencias clave en los currículos y en la evaluación ha tenido como consecuencia reformas educativas en distintos países, como en Francia (MINISTERE, 2006) o España (LOE, 2006), en las que se ha llevado a cabo esta incorporación de las competencias básicas (Roca y Sánchez, en prensa). Además, en los diversos programas de evaluación de distintos países y de la propia UE se han venido incorporando a las áreas tradicionales de estudio, transformadas ahora en competencias básicas (lectura, matemáticas, ciencias), nuevas competencias básicas: la ciudadanía (evaluada por ICCS de IEA, con módulo específico europeo, y la participación de los siguientes países latinoamericanos: Chile, Colombia, Guatemala, México, Paraguay y República Dominicana), las tecnologías de la información y la comunicación (PISA-ERA-Electronic Reading Assessment), o la competencia de los alumnos para aprender a aprender (UE).

El enfoque curricular (basado en los diversos tipos de contenidos, comunes en los currículos de los países participantes) se mantiene en importantes estudios internacionales: PIRLS (donde participaron Argentina y Colombia en la edición del 2001, y España en la edición 2006) y TIMSS de la IEA (que incluye en el 2007 a Colombia y España) o los estudios del LLECE. Este enfoque permite orientar el estudio hacia los aprendizajes, los conocimientos y las destrezas adquiridos por los alumnos en las áreas examinadas. La principal dificultad de este enfoque reside en que, en un estudio internacional, es necesario delimitar los aspectos comunes de los distintos currículos y limitar las pruebas a esos aspectos. En muchos casos, objetivos, criterios y estándares de los distintos currículos no facilitan esta tarea. Por esta razón se han venido usando preferentemente las áreas instrumentales y, dentro de ellas, dominios específicos de la comprensión lectora o las matemáticas. El lector encontrará una detallada reflexión sobre este asunto en el capítulo de Margarita Poggi en este libro.

Estos inconvenientes no son muy diferentes en el enfoque de competencias básicas si, como señalan distintos expertos, no existe un suficiente consenso teórico sobre la definición de las competencias básicas o no hay acuerdo generalizado sobre las dimensiones de las competencias que se deben considerar y evaluar. Los defensores del modelo basado en competencias subrayan que, en las distintas unidades de evaluación, reproducidas mediante los denominados estímulos sobre los que se organizan las cuestiones concretas, se enfrenta a los alumnos a situaciones reales ante las que, para resolverlas correctamente, los alumnos deben utilizar los distintos conocimientos, destrezas y actitudes adquiridas en el conjunto de las áreas de los currículos, y no solo en aquellas áreas más próximas a las competencias evaluadas.

Factores asociados

Los estudios internacionales han permitido constatar que el rendimiento educativo de los alumnos está relacionado con distintos factores, entre los que se señalan como fundamentales el estatus social, económico y cultural de las familias, los relacionados con el trabajo y el clima en el aula, o la organización de las escuelas. Por esta razón, dichos estudios elaboran diferentes índices estadísticos que pretenden ofrecer información sobre la relación entre los resultados y el factor concreto del que se trate. Uno de estos índices, hoy utilizado en prácticamente todos los estudios internacionales, es el denominado índice de estatus social, económico y cultural (ESEC), calculado a partir de las respuestas de los alumnos y de los cuadernillos de profesores, directivos y familias.

La elaboración de este índice para el conjunto de países y regiones participantes en los distintos estudios suele tener en cuenta diferentes componentes: el nivel de educación o formación alcanzado por los padres, su situación laboral y su profesión y el nivel de recursos educativos en el hogar (espacio para estudiar, conexión a Internet, libros de estudio y de literatura, etc.).

En diferentes estudios, este índice proporciona un valor para cada país o región, escuela o alumno. Ello permite considerar la relación entre el rendimiento y el estatus socioeconómico y cultural de familias y escuelas dentro de cada país o región y comparar su comportamiento con los de otros. Combinando las puntuaciones promedio y el índice se pueden construir líneas de regresión (gradiente) que resumen la relación entre las diferentes puntuaciones del rendimiento y los distintos valores ESEC de escuelas y alumnos de un país. Estos gradientes ponen de manifiesto que, en muchos casos, a similares niveles de bienestar social, económico y cultural corresponden resultados diferentes, indudablemente justificados por el trabajo en el aula y en la escuela. También muestran la conveniencia de mejorar los rendimientos de *todos* los alumnos (excelencia) y disminuir, al tiempo, las diferencias (equidad), así como la necesidad de proporcionar en la escuela entornos educativos que compensen entornos familiares y escolares desfavorecidos (MEC, Instituto de Evaluación, 2007c).

De interés particular ha resultado el índice sobre el clima escolar calculado por el SERCE (LLECE, 2008), con el que se pone de manifiesto la importancia que tiene mejorar la organización y el funcionamiento de las escuelas, el trabajo de los profesores y su relación con los alumnos, pues las variables consideradas en estos factores relacionados con el clima escolar son más determinantes cuanto más pequeños son los niños. En la consideración de los factores asociados conviene tener presente la rica información que ofrecen los distintos indicadores de los sistemas educativos. En este sentido, se remite al lector al capítulo de Margarita Poggi en este libro.

Análisis y presentación de los resultados: la potencial utilidad de los informes internacionales

En la presentación de los resultados de la mayoría de evaluaciones internacionales no se utiliza la metodología conocida como "teoría clásica de los tests", según la cual las puntuaciones obtenidas por los alumnos se expresan en porcentajes de aciertos o puntos, sobre el total que puede obtenerse en la prueba. En su lugar, se utiliza el método conocido como "teoría de respuesta al ítem" (TRI), que permite situar el resultado o nivel de competencia de cada alumno en una escala común, sean cual sean los bloques de la prueba. Los resultados globales alcanzados por los alumnos se suelen expresar en una escala continua en la que se hace equivaler a 500 puntos el promedio de

las puntuaciones medias de los participantes en el estudio. Este p*romedio de 500* es, por tanto, un *índice* de los *resultados medios* de los países y alumnos participantes.

Hay que tener en cuenta, por tanto, que estas puntuaciones no son *calificaciones* como las que otorga habitualmente un profesor a sus alumnos cuando evalúa el resultado de su aprendizaje. El promedio de 500 puntos es un número índice que facilita la comparación entre países, número que nada tiene que ver con una supuesta calificación sobre un criterio de excelencia (no cabe en modo alguno establecer equivalencias con el valor de los puntos en una escala de 1 a 10).

Pero el resultado promedio global, siendo útil, pues resume en una cifra por país y en una escala al conjunto de los países, no deja de ser una parte pequeña de la rica y compleja información que las evaluaciones internacionales ofrecen. Hay que resaltar la enorme utilidad de la presentación de resultados para cada área o competencia evaluada, según distintos niveles de rendimiento de los alumnos en dicha área o competencia. Esta presentación la realizan prácticamente todos los estudios (PIRLS, TIMSS, PISA, SERCE…) y permite verificar, por un lado, qué porcentaje de alumnos se encuentra en cada uno de los niveles de rendimiento reconocidos. Por otro lado, permite describir los conocimientos, habilidades o competencias que caracterizan a cada uno de los niveles. El valor relativo de estos niveles y el marco de comparación que ofrecen es un instrumento muy valioso para los sistemas educativos, que encuentran contrastados los rendimientos de sus alumnos con estas referencias de comparación internacional. Es cierto que desde el punto de vista metodológico no es sencillo definir niveles de rendimiento, por lo que no es prudente comparar los que se establecen en un estudio con los de los demás, a no ser que se adopten las cautelas y las consideraciones técnicas pertinentes.

Una segunda información de inestimable valor que ofrecen los estudios internacionales es la relativa a los resultados en cada uno de los principales aspectos o dimensiones del área o competencia evaluada. Es posible así conocer los resultados de los alumnos según los distintos procesos cognitivos propios de un área o competencia (por ejemplo, en la comprensión lectora, se encuentran valoraciones de la capacidad de los alumnos para aproximarse al texto e identificar, organizar, integrar, sintetizar, reflexionar, valorar, transferir y aplicar la información) o según los bloques de contenidos (por ejemplo, los números y operaciones, la estimación y cálculo de magnitudes, las formas y la geometría, o el tratamiento de la información, el azar y la probabilidad en la competencia matemática).

Por otra parte, todos los estudios ofrecen diferentes medidas de las variaciones de los resultados que presentan los alumnos de un mismo país o región y, por tanto, de la capacidad de los sistemas educativos de ofrecer a todos los alumnos, o a un número muy elevado de ellos, niveles elevados y equitativos de rendimiento. Un resultado promedio elevado puede deberse a distribuciones muy desiguales entre las puntuaciones obtenidas por los alumnos de modo que haya, junto a unos pocos con resultados excelentes, otros con resultados muy pobres. Un sistema educativo solo puede ser de calidad si es capaz de ofrecer al mismo tiempo elevado rendimiento y equidad o, dicho de otro modo, elevados resultados para todos o una inmensa mayoría de sus alumnos. Los estudios internacionales ofrecen en este sentido una información y una referencia en la comparación muy estimables.

Finalmente, como se ha señalado antes, los estudios internacionales ofrecen también hoy los resultados de los alumnos en relación con diferentes factores asociados, entre los que cabe destacar los índices sociales, económicos y culturales, en la casi totalidad de los estudios actuales, y otros índices como el clima escolar (LLECE, 2008).

Todos los estudios ponen de manifiesto el peso relativamente importante de estos factores en el rendimiento de los alumnos. Pero en todos los estudios se puede observar también que alumnos y escuelas sometidos a factores contextuales similares pueden presentar resultados muy diversos. O, dicho de otro modo, que es fundamental el trabajo de profesores y alumnos, que los procesos de enseñanza y aprendizaje que se desarrollan en el aula y en el trabajo cotidiano terminan modificando, a veces de modo muy positivo, factores de contexto y de entorno desfavorables.

LOS ASPECTOS MENOS SATISFACTORIAMENTE VALORADOS POR LOS ESTUDIOS INTERNACIONALES: PROCESOS EDUCATIVOS, FUNCIONAMIENTO DE LAS ESCUELAS, TRABAJO DEL PROFESORADO

Los estudios internacionales de evaluación tienen como propósito fundamental evaluar el funcionamiento de los sistemas educativos. Generalmente han tratado de cumplir este cometido mediante el análisis del rendimiento alcanzado por una muestra significativa de alumnos del país o región considerada, aspecto fundamental, que además ofrece la ventaja de que es cuantificable con relativa precisión y sencillez. Pero, al mismo tiempo, los propios estudios no han dejado de insistir de modo reiterado en que, siendo el rendimiento de los alumnos un indicador valioso del funcionamiento del sistema educativo, hay otros muchos aspectos que deben ser tenidos en cuenta: cómo se producen los aprendizajes (los procesos educativos), cómo están preparados (formación), trabajan y colaboran los profesores para que el rendimiento de sus alumnos sea el máximo posible, cómo deben organizarse y funcionar las escuelas para lograr el mismo fin, cuáles son las circunstancias y los contextos de los alumnos y de las propias escuelas que explican los rendimientos medidos, cómo influye la implicación de las familias y de la sociedad. Todas estas cuestiones han sido objeto de preocupación y han ido incorporándose con mayor o menor intensidad y acierto a la mayoría de los recientes estudios, junto con la medida del rendimiento de los alumnos.

Se ha señalado insistentemente la importancia de medir los procesos educativos. Sin embargo, mientras la valoración de resultados de los alumnos ha progresado de modo notable y satisfactorio, la medición de los aprendizajes y de los procesos educativos que tiene lugar entre el maestro y sus alumnos sigue siendo compleja y poco exitosa. No es posible la observación directa de los procesos a gran escala y de modo confiable y no es fácil cuantificar su eficacia de modo objetivo, dos aspectos fundamentales en la evaluación que no se logran resolver de modo satisfactorio en la valoración de los procesos educativos. La información principal sobre los procesos que obtienen los estudios internacionales proviene de los cuestionarios que responden alumnos, profesores y equipos directivos. Hasta la fecha, esta información sigue siendo escasa en la mayoría de los estudios. No obstante, estudios como TALIS de la OCDE abren nuevas posibilidades sobre la evaluación de los procesos, que es de esperar sean fructíferas.

Por otra parte, los estudios internacionales proporcionan información sobre la variabilidad de los resultados entre escuelas, lo cual permite valorar hasta qué punto las diferencias de los resultados de los alumnos dependen de su escolarización en uno u otro centro. La influencia de las escuelas sobre los resultados de los alumnos puede deberse a distintas razones: el currículo real impartido, el clima escolar, el ambiente de estudio o el origen social conjunto de los alumnos matriculados. Pero los estilos propios de las escuelas benefician o perjudican en mayor o menor medida a sus alumnos. Como se ha señalado, los centros educativos pueden ejercer un notable efecto sobre los resultados de sus alumnos: las escuelas con buenos resultados globales mejoran los resultados de los alumnos de menor rendimiento.

La mayoría de los estudios señalan la conveniencia de que las escuelas puedan disponer de autonomía para adoptar las acciones necesarias, de los recursos suficientes y de planes de trabajo encaminados a la consecución de los objetivos y la evaluación de sus procesos, sus contextos y sus resultados. Desde esta perspectiva, la autonomía de los centros aparece como un factor que puede favorecer de modo notable la mejora de los resultados educativos de sus alumnos.

Un aspecto que está en el centro de las preocupaciones evaluadoras y constituye un reto fundamental en la evaluación, en general, y en la de las escuelas, en particular, es el valor añadido que estas son capaces de ofrecer a sus alumnos. Existe un proyecto de la OCDE que pretende abordar precisamente esta cuestión, pero que se encuentra todavía en una fase inicial de valoración de la situación actual en este campo.

Finalmente, como se ha señalado, la mayoría de los estudios muestran la importancia del profesorado en el rendimiento de los alumnos, pero en estos estudios de evaluación casi son inexistentes los apartados dedicados a valorar la formación, inicial y permanente, el reconocimiento, la exigencia, la promoción y la gratificación del profesorado y sus efectos en la mejora de la calidad de los sistemas educativos. Proyectos como TALIS (OCDE) y TEDS-M (Formación Inicial del Profesorado en Matemáticas de la IEA, donde participan España y Chile) pueden abrir un camino que es imprescindible transitar en los próximos años.

CONCLUSIÓN

Los estudios internacionales de evaluación han adquirido, sobre todo desde la aparición de PISA, un impacto mediático considerable en la mayoría de los países. Si bien la presentación que la prensa hace suele quedarse en el impacto de las clasificaciones internacionales de los países según los resultados obtenidos por sus alumnos y, en muchas ocasiones, estas presentaciones han sido utilizadas para generalizar las críticas a los responsables y agentes educativos, lo cierto es que las repercusiones a medio plazo para los sistemas educativos están siendo crecientemente importantes. Por un lado, los estudios ofrecen una información imprescindible sobre la situación de la educación y, por tanto, facilitan la toma de decisiones por parte de los responsables de las políticas educativas para mejorarlos. Véanse en este sentido las reflexiones que se ofrecen en los dos últimos capítulos del presente volumen.

Pero no solo facilitan la toma de decisiones de los responsables. La información que ofrecen los estudios, cuando se propicia y consiguen las condiciones idóneas para su uso, es de la mayor utilidad para la comunidad educativa. Los alumnos y sus familias tienen una referencia nacional e internacional de los aspectos más relevantes de los aprendizajes y sus niveles. Los profesores reciben información sobre las fortalezas y debilidades de su trabajo y del de sus alumnos, y las circunstancias que mejor las explican. Los equipos directivos encuentran referencia a los aspectos organizativos y a los recursos que resultan más eficaces en el rendimiento de los alumnos. Finalmente, las sociedades pueden conocer mejor el funcionamiento de sus sistemas educativos y, como consecuencia, pueden valorar de modo más ajustado los aciertos, las deficiencias y las dificultades, pueden exigir mejores medidas y mayor esfuerzo a todos los responsables educativos y pueden implicarse más en la asignación de recursos que una educación de calidad para todos reclama y en el apoyo a sus jóvenes y a sus maestros.

Los resultados de las evaluaciones internacionales pueden ser un arma arrojadiza, pueden ser utilizadas de modo inadecuado o pueden producir un efecto contraproducente como resultado

de una inadecuada comprensión de sus alcances y limitaciones; esto ocurre, por ejemplo, cuando se utilizan exclusivamente los resultados promedio de un país y el lugar que ocupa en el conjunto para valorar su sistema educativo. Pero este es un riesgo que prácticamente desaparece si se acierta en la difusión y el buen uso de los resultados de las evaluaciones y que queda sobradamente compensado por el hecho de que las evaluaciones internacionales pueden ser, sobre todo, un valioso e indispensable instrumento para conocer y poder adoptar las medidas necesarias para la mejora de la educación.

Las evaluaciones regionales y nacionales en América Latina

Guillermo Ferrer

INTRODUCCIÓN

La construcción de un sistema nacional de evaluación es un emprendimiento complejo, y de esto pueden dar cuenta casi todos los países de la región. Más de dos décadas han pasado desde las primeras experiencias de evaluación estandarizada a gran escala en América Latina y, aunque mucho se ha avanzado en la consolidación, en el mejoramiento técnico y en la legitimación de los proyectos y programas nacionales de evaluación educativa, queda todavía un largo camino por andar[1].

En una región donde las condiciones para el trabajo docente son precarias, con escuelas que albergan poblaciones de alumnos cada vez más numerosas y diversas, y donde el Estado no suele ser el mejor ejemplo de responsabilidad y coherencia en el diseño e implementación de políticas, la evaluación externa y estandarizada ha debido realizar grandes esfuerzos por abrirse paso y hacerse lugar en la compleja y abultada agenda de prioridades en el terreno de la educación pública. Ha debido también afinar la capacidad de diálogo con las tradiciones pedagógicas e ideológicas locales, casi siempre enfrentadas con la idea de la evaluación externa y objetiva, y temerosas de los usos punitivos que de ella puedan hacerse.

Cuando hablamos de sistemas o programas de evaluación educativa en América Latina, hablamos principalmente de *equipos*: grupos de profesionales que, desde algún órgano de gobierno o institución académica convocada oficialmente, asumieron el desafío de diseñar e implementar operativos de evaluación a gran escala. A veces lo hicieron en el marco de una reforma integral del sistema educativo, y como parte de un esquema renovado de gestión pedagógica y administrativa, otras como proyectos marginalmente vinculados a las políticas del sector. En ambos casos, sin embargo, muchos de estos equipos han logrado sostenerse en el tiempo, transitando gobiernos de diferentes banderas, acumulando experiencia técnica y política, y aportando progresivamente a la discusión sobre los procesos y resultados de las acciones en materia de educación pública. Otros equipos, aunque no lograron permanecer en sus funciones al cabo de las primeras experien-

[1] Todas las informaciones presentadas en este capítulo han sido recabadas, documentadas y difundidas en múltiples publicaciones por el Grupo de Trabajo sobre Estándares y Evaluación (GTEE) del PREAL, del que este autor es miembro investigador desde 1999. El texto aquí presentado, aunque de producción enteramente original para esta publicación, sintetiza datos y reflexiones de varios estudios del GTEE que son citados a lo largo del capítulo. Para obtener informaciones más precisas sobre la evaluación en cada país de la región, se invita al lector a consultar las fuentes de referencia.

cias, sí dejaron un antecedente técnico y cierta capacidad operativa que tiempo después permitió reanudar acciones de evaluación bajo nuevas administraciones del sector.

Valga esta descripción para comprender que un auténtico *sistema* de evaluación educativa en nuestros países no se funda de la noche a la mañana. Depende mayormente del desarrollo de capacidades técnico-profesionales de personas que son convocadas y deciden dedicarse con entrega y compromiso a una tarea mayúscula en un campo donde el conocimiento previo escasea y donde los condicionantes políticos e ideológicos abundan. En consecuencia, la consolidación de un sistema de evaluación obliga a una fuerte interdependencia entre las agendas de gobierno referidas a evaluación y las personas que pueden materializarla con continuidad y solidez.

La evolución de esa relación en los diferentes países ha permitido, en algunos casos, que los proyectos primigenios y acotados de medición hayan devenido en sistemas más integrales y articulados de evaluación educativa. En otros casos, aun cuando se hayan mantenido los equipos y la regularidad en las aplicaciones, e inclusive logrado el crecimiento en la cobertura poblacional de los operativos, los esfuerzos no alcanzan para constituir auténticos sistemas de evaluación; antes bien todavía pueden considerarse proyectos de medición con escasa o nula articulación con los demás componentes de la gestión.

Así es como, a lo largo de varios años de reforma educativa en la región, con altos grados de rotación de autoridades y recambios políticos, con inestabilidad en las condiciones para el mantenimiento de elencos técnicos, con demandas de producción de información no siempre bien definidas ni ajustadas a un esquema coherente y sistémico de política educativa, con errores y aciertos, pero con decisión y voluntad de superación, se fueron conformando y afianzando los sistemas nacionales y subnacionales de evaluación.

CARACTERÍSTICAS DE LOS SISTEMAS DE EVALUACIÓN

Todos estos sistemas comparten algunas características comunes y también exhiben rasgos que los diferencian, tanto en su diseño y constitución actual como en la evolución que han experimentado en los últimos años. Entre los múltiples aspectos que pueden considerarse para describir estos sistemas en la actualidad, seleccionamos aquí los referidos a sus fines y propósitos, al marco institucional para su funcionamiento, a las principales características de los operativos e instrumentos de medición y a la forma como los datos obtenidos se traducen en información útil para las diferentes audiencias[2].

Fines y propósitos de la evaluación

En algunos países, la constitución del sistema nacional de evaluación tuvo fines muy específicos, como por ejemplo los de regular la admisión a la educación superior con pruebas aplicadas a los alumnos que egresan de educación media, o los de promover la elección de escuelas por parte de las familias sobre la base de información objetiva y públicamente accesible. En la mayoría de los casos nacionales, sin embargo, y particularmente los desarrollados a partir de la década de los noventa, los propósitos de los sistemas de evaluación se expresaron en términos mucho más amplios: proveer información para la toma de decisiones de política y gestión en los niveles central y local, así como para el mejoramiento continuo del currículo y de los procesos pedagógicos. Apa-

[2] Categorías temáticas extraídas de Ferrer (2006a).

recen frecuentemente expresados también los propósitos de estimular la opinión y participación de la sociedad civil acerca de la calidad de la educación pública, y los de promover "una cultura de evaluación" dentro del sistema educativo.

En ambos casos, esos propósitos han sufrido ajustes y redefiniciones en función de la experiencia y de las demandas concretas, y siempre cambiantes, de datos e información por parte del sistema. Los programas de evaluación, o "servicios de medición", que iniciaron acciones con propósitos más acotados y enfocados en la responsabilización (*accountability*, altas implicancias), desarrollaron con el tiempo mayor capacidad para proveer información de utilidad pedagógica –central y local–, atendiendo principalmente a la necesidad de revelar algunos de los procesos y factores que pueden explicar los logros o insuficiencias en el rendimiento del alumnado. Por su parte, los sistemas que plantearon metas muy amplias, y por lo general inabarcables en la práctica, debieron priorizar algunas de ellas y seleccionar los instrumentos y tipos de operativos que mejor permitieran cumplirlas.

Para todos, sin embargo, sigue siendo un desafío poder abarcar todos los aspectos del sistema que influyen en la calidad educativa de un país y no solamente los resultados de aprendizaje de los estudiantes. El desafío implica, por un lado, conceptualizar la complejidad del sistema, sus componentes y las formas como estos interactúan en los procesos educativos. Por ejemplo, la formación y el desempeño de los docentes, los recursos pedagógicos disponibles en las escuelas o algunos factores extraescolares tales como el capital sociocultural de las familias. Por otro lado, implica desarrollar las capacidades técnicas y operativas necesarias para el diseño de modelos e instrumentos que permitan medir todas esas variables. En este sentido, lenta pero sostenidamente, y conforme se desarrollan las capacidades de los equipos técnicos, los programas van dejando de ser proyectos o experimentos de medición para convertirse en sistemas más integrales de evaluación.

Marcos institucionales

Así como cambian o evolucionan los fines de la evaluación, también van cambiando los marcos institucionales en los que operan los programas. Durante las dos últimas décadas, los países han ensayado diferentes marcos institucionales para la administración de los sistemas de evaluación educativa. En general, puede afirmarse que, a pesar de los cambios y de algunos casos de fuerte inestabilidad política, ha habido una tendencia a sostener los sistemas de evaluación nacionales y a reinstaurarlos o recomponerlos cuando sus actividades fueron discontinuadas.

En varios países, los arreglos institucionales más estables resultaron ser los que se instalaron fuera de la estructura orgánica de los Ministerios de Educación. Estos sistemas, generalmente denominados "institutos", poseen mayor autonomía administrativa y técnica que otros órganos de línea y suelen llevar a cabo sus tareas de medición y devolución de resultados con más agilidad y consistencia que los sistemas dependientes de los Ministerios. En otros casos se ha dado una evolución inversa, en que sistemas o proyectos de medición anteriormente administrados por instituciones independientes se encuentran hoy administrados desde los Ministerios de Educación, con el propósito de acercarlos a los demás órganos de gestión y así promover el trabajo conjunto.

En cualquiera de estas situaciones de cambio puede advertirse, o por lo menos interpretarse, que los sistemas de evaluación van cobrando protagonismo y que, lejos de desvanecerse, son siempre objeto de debate político y administrativo para hacerlos más útiles y eficientes. Es constante,

a la vez, la actualización normativa que va haciendo de estos sistemas órganos más estables y más imbricados en la burocracia del sector.

Esta evolución en el "estatus" de los sistemas de evaluación ha significado mucho más que estabilidad en sus acciones. Ha permitido a los elencos técnicos, principalmente, acceder a más y mejores oportunidades de intercambio y capacitación con pares extranjeros y organizaciones internacionales especializadas. Organismos y agencias tales como la OEI, UNESCO-OREALC, el Banco Mundial y PREAL han facilitado estos intercambios y aprendizaje mediante seminarios, talleres, cursos, publicaciones y también, como se tratará más adelante, la participación en proyectos internacionales de evaluación comparada. En muchos sentidos, estas oportunidades de especialización y desarrollo han sido más frecuentes, sistemáticas y enfocadas que las disponibles para otros órganos o equipos de gestión de los Ministerios nacionales de Educación.

Conforme se fue ganando experiencia, accediendo a oportunidades de capacitación y re-evaluando las necesidades de información de los sistemas educativos, fueron cambiando también los modelos e instrumentos de medición utilizados. Estos cambios también han estado vinculados a los procesos de revisión y actualización de los currículos nacionales, en la medida en que la evaluación estandarizada precisa, y también contribuye a construir, referentes más sólidos y operacionales para comunicar eficazmente las expectativas de aprendizaje y sus alcances.

Tipos de instrumentos

Uno de los cambios más significativos, y más generalizados en la región, fue la transición de *pruebas referidas a normas* a *pruebas referidas a criterios*. Mientras que las primeras servían el propósito principal y casi exclusivo de comparar el rendimiento entre diferentes poblaciones de un sistema, las últimas permiten contrastar el rendimiento de cada alumno o subpoblación contra un conjunto de expectativas comunes de aprendizaje. De esta manera, las evaluaciones nacionales de casi todos los países hoy pueden decir algo más que "X obtiene mejores resultados que Y"; pueden, con mayor o menor grado de especificidad, reportar qué saben diferentes grupos de estudiantes respecto a un conjunto significativo de saberes planteados por el currículo nacional. La diferencia entre un modelo y otro no es meramente metodológica, sino que atiende a una necesidad muy concreta de información para el mejoramiento de los procesos curriculares y pedagógicos del sistema[3].

Simultáneamente, los avances recientes en la teoría y los modelos psicométricos han permitido el desarrollo de instrumentos que pueden medir y reportar los aprendizajes no solamente en función de *criterios*, sino también por *niveles de desempeño*, que indican grados de profundidad en el dominio de los saberes evaluados. Dicho en otras palabras, permiten medir no solamente "qué saben" los estudiantes, sino también "cuán bien lo saben y cómo lo demuestran" a partir de *indicadores* muy precisos que dan cuenta del nivel de logro alcanzado.

La incursión de estos modelos ha estado promovida no solamente por los desarrollos de nuevas teorías y metodologías de medición, sino también por la creciente difusión del concepto de *estándares*[4] y sus implicancias para la actualización en las formas de prescribir, comprender, socializar

[3] Descripciones más detalladas de los enfoques normativos, criterial y de progreso pueden consultarse en Ravela (ed.) (2001) y Ravela (2006).

[4] En Ferrer (2006 b) se encuentra un tratado exhaustivo del concepto de *estándares* como instrumentos

y evaluar el currículo y los aprendizajes. Si bien no es posible ahondar en este tema aquí, valga su mención para destacar la importante y necesaria relación entre una buena medición estandarizada –de utilidad pedagógica en todos los niveles de un sistema– y el diseño y utilización de referentes curriculares complejos y a la vez adecuadamente operacionalizados. En su capítulo, Elena Martín analiza este tema más a fondo. Ya son varios los países que, aunque con diferentes abordajes y formas de resolución técnica, han decidido revisar sus currículos y plantear *estándares de contenido* y *desempeño* que permitan una mejor medición y evaluación de los aprendizajes, tanto en forma estandarizada a gran escala como en los procesos pedagógicos de aula[5].

Estos avances en los modelos de evaluación utilizados en América Latina también han sido facilitados por el creciente protagonismo y difusión de las pruebas internacionales comparadas de aprendizaje[6], especialmente las lideradas por la OCDE (Organización para la Cooperación y el Desarrollo Económico), por la IEA (International Association for the Evaluation of Educational Achievement) y por UNESCO-OREALC a través de su Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE). En su capítulo en esta obra, Enrique Roca analiza estos programas con detenimiento.

El impacto de estos proyectos internacionales en los sistemas locales tiene dos aristas. Por un lado, la participación de los países de la región en estas pruebas ha facilitado el acceso de los equipos técnicos a las nuevas metodologías y recursos técnicos de complejidad en materia de medición. La incorporación del modelo TRI (teoría de respuesta al ítem), la definición de niveles de desempeño sobre base empírica, el establecimiento de líneas de corte, la exhaustiva descripción conceptual y operacionalización de los saberes disciplinares que se han de evaluar, y los procedimientos para optimizar la confiabilidad en la corrección de preguntas abiertas son apenas algunos ejemplos del aporte teórico y metodológico de los proyectos internacionales. La capacidad creada a partir de una experiencia de este tipo sería difícil de igualar si se intentara mediante cursos o talleres ad hoc, en forma dispersa e individual para cada equipo nacional. Esta reflexión, naturalmente, cobra mayor pertinencia para países con poca tradición o corta experiencia en evaluación criterial estandarizada.

Por otro lado, las pruebas internacionales, y PISA-OCDE en particular, han contribuido significativamente al desarrollo de un lenguaje común respecto a varios aspectos del currículo y la evaluación. El concepto de alfabetización científica o comprensión lectora como competencias para la vida, la forma en que diferentes ejes de una misma disciplina se abordan integrada y sistemáticamente desde su dimensión procedimental, o la posibilidad de establecer niveles de desempeño crecientes –e inclusivos– de complejidad, son también ejemplos de los tipos de conceptos y terminología que en el ámbito educativo suelen interpretarse de maneras muy diferentes cuando no se trabajan dentro de un mismo marco de referencia. Las pruebas internacionales comparadas permiten, al menos durante su desarrollo y posterior difusión de resultados, centrar la mirada de

técnicos y de política curricular.

[5] En Ferrer (en elaboración) puede consultarse el proceso de elaboración e implementación de estándares en cinco países de la región.

[6] Ver el Anexo a este capítulo con una síntesis de las pruebas internacionales en las que han participado países iberoamericanos. Mayores descripciones y datos acerca de estas pruebas internacionales, así como de los países de la región que han participado en ellas, pueden consultarse en Ferrer y Arregui (2003), en Ravela (2006) y en los sitios web de las organizaciones mencionadas.

actores de muy diversos contextos en aspectos y problemas comunes del currículo, el aprendizaje y la evaluación.

En resumen, la acumulación de experiencia local y la aportada por los proyectos de evaluación internacional comparada han hecho posible que hoy nuestros países exhiban una variedad de instrumentos de medición más actualizada, de mayor cobertura y profundidad curricular y, por lo tanto, de mayor pertinencia para los destinatarios de la información. Comparando la situación actual con los modelos de evaluación de una década atrás, cuando las pruebas eran principalmente referidas a normas y los ítems o actividades se planteaban casi exclusivamente en formato de opción múltiple, se puede afirmar que los avances metodológicos en evaluación han sido significativos.

Formas de reporte, devolución y uso de los resultados

La actualización en los modelos y tipos de instrumentos impacta directamente sobre la forma en que se reportan y comunican los resultados de las mediciones. En este terreno también se ven adelantos interesantes, en la medida en que los sistemas de evaluación van afinando sus estrategias y diseños para llegar más "amigablemente" a los usuarios pretendidos. En el capítulo de Patricia McLauchlan se desarrolla este tema con más detalle.

Tanto la evaluación por criterios como los modelos que permiten establecer niveles de desempeño habilitan el diseño de reportes mucho más interesantes y completos que el simple reporte de porcentaje de respuestas correctas. La tendencia en la región es la de presentar informes que integran, cada vez con más solvencia gráfica y textual, datos numéricos con buenas descripciones verbales del referente evaluado. En un caso típico, se presentará un gráfico de la escala utilizada para ordenar los resultados y se dividirá esa escala en tramos de puntaje; a su vez, cada tramo estará acompañado por una descripción de los tipos de saberes demostrados por los alumnos o poblaciones que allí se ubicaron según los puntajes obtenidos. Paralelamente se presentará los porcentajes de alumnos de diferentes contextos que alcanzaron cada uno de esos niveles que los tramos definen.

Este tipo de reportes facilita la comprensión de los resultados y la reflexión sobre cómo mejorar los procesos curriculares y didácticos. Para audiencias políticas, la prensa y la opinión pública, la comprensión de estos informes puede promover reflexiones más sustantivas, más vinculadas al contenido de la educación y sus problemas, y no enfocadas exclusivamente en las cifras alarmantes de los resultados, por verdaderas que sean.

Desde una perspectiva pedagógica más amplia, la medición y los informes de resultados basados en niveles progresivos de desempeño guardan el potencial de cambiar algunas concepciones clave respecto a la evaluación, al aprendizaje mismo y a la enseñanza. El principal aporte en este sentido sería el de superar la visión de que los estudiantes saben o no saben, son buenos o malos, sino que todos se encuentran en algún punto del recorrido y que, acciones adecuadas mediando, pueden alcanzar las metas que el sistema les promete. Aunque esto parezca obvio, trillado quizá, lo cierto es que en las escuelas circula un alarmante determinismo respecto a qué pueden aprender o no algunos estudiantes, y cuando la evaluación no es sustantiva y se centra puramente en las comparaciones, no hace más que reforzar esas concepciones.

El diseño y utilización de modelos más actualizados de evaluación puede, por el contrario, ayudar a concebir el aprendizaje como un proceso continuo, auténticamente constructivo, pero en el

que el logro de las metas objetivamente descritas y evaluadas importan tanto como el recorrido. En otras palabras, se sostienen metas comunes para todos los estudiantes de un sistema, a la vez que se diversifican y adecúan los recursos y las estrategias pedagógicas para ayudar a que todos alcancen los objetivos.

Importa aquí destacar que, en la mayor parte de los países de nuestra región, estos modelos y las concepciones pedagógicas que los sustentan han sido promovidos principalmente por los sistemas de evaluación. Esto no es un dato menor, considerando que unos años atrás los equipos de evaluación tenían escasa o nula injerencia en materia de currículo y pedagogía: se limitaban a tomar del currículo oficial aquello que fuera "medible" con los instrumentos disponibles, para luego informar sobre qué grupos de alumnos dominaban o no ciertos saberes.

Actualmente, son cada vez más los países de América Latina que, a partir de estos modelos, van integrando el trabajo de los equipos de evaluación con los de currículo. Esto demuestra en cierta medida que existe un mayor entendimiento de que no se puede evaluar bien sin buenos referentes de currículo y que es estéril la labor de desarrollo curricular si no es acompañada por un sistema de evaluación idóneo que pueda dar cuenta de los logros progresivos de los estudiantes[7]. En su capítulo, Elena Martín analiza con más detalle las complejas relaciones entre evaluación estandarizada y currículo.

Otro aspecto de la forma en que se reportan los resultados de las evaluaciones es el de la información respecto a la equidad que logran los sistemas educativos. Para que los datos de rendimiento tengan sentido y relevancia para la política educativa, deben estar ligados a los datos que el sistema recoge acerca del contexto socioeconómico y cultural de los estudiantes evaluados, y complementados con diversos análisis inferenciales que expliquen las diferencias de resultados entre poblaciones o subsistemas. Esto permite contemplar los entornos y los tipos de estudiantes que las escuelas atienden, y así contextualizar los resultados de las evaluaciones.

Estos análisis resultan útiles cuando permiten no solamente clasificar los logros de estudiantes de diferentes grupos socioeconómicos, sino cuando además permiten medir el aporte educacional, o *valor agregado*, que realizan las escuelas en cada uno de esos contextos. Por ejemplo, no resulta muy útil comprobar año tras año que los estudiantes de escuelas rurales obtienen resultados más bajos que los de escuelas urbanas. Ese dato "fotográfico" poco aporta para entender qué ocurre educacionalmente en ambos tipos de escuelas. Sin embargo, si comparamos año tras año cuánto evolucionó cada escuela en su contexto, es posible descubrir que muchas escuelas rurales logran más avances que las urbanas, aunque en términos absolutos sus resultados sigan siendo inferiores. Este es un dato sumamente importante para los sistemas educativos, o debiera serlo al menos, ya que da en el centro de la problemática de la equidad. Los programas de evaluación entonces tienen mucho para aportar con este tipo de informaciones y, aunque lentamente, cada vez son más los países que lo están haciendo.

Disponer de buenos datos, sin embargo, no es suficiente para lograr impacto en las decisiones de política y en los procesos pedagógicos del sistema. Para ello, es necesario primero transformar los

[7] Para comprender mejor este acercamiento entre unidades de currículo y evaluación se recomienda la lectura del caso chileno en Ferrer (en elaboración) y de los modelos integrados de currículo y evaluación desarrollados por el ACER (Australian Council for Educational Research).

datos en *informaciones* sustantivas. Luego, encontrar los mejores canales y formatos de entrega para hacer llegar la información a las audiencias previstas.

La primera condición se va cumpliendo cada vez mejor, como se explicó anteriormente, mediante el diseño de informes que resultan más sustantivos y amigables al lector que los que se encontraban años atrás. No solo se van mejorando los reportes, sino que además se van diferenciando en formato y contenido según las necesidades de diferentes públicos y usuarios de la información. Por ejemplo, el nivel de agregación de datos que recibe una funcionaria en la gestión federal de un sistema es muy diferente del que se le hace llegar a un docente para su trabajo en el aula o a un supervisor de escuelas en un distrito comunal. Los avances en este sentido han implicado un cambio de concepción, a veces con bastante renuencia, respecto al papel que juegan los evaluadores en un sistema. ¿Deben los evaluadores ser también comunicadores? ¿O deben concentrar sus energías en la producción y entrega oportuna de datos válidos y confiables? Sea por opción, o porque las circunstancias lo exigen, lo cierto es que los evaluadores asumen cada vez con mayor frecuencia el compromiso de transformar datos en informaciones para lograr una mejor comprensión de los resultados de las evaluaciones.

La segunda condición, referida a los canales y formatos de entrega, también viene siendo objeto de mejora en los últimos años. Ciertamente los países que aplican pruebas censales, y en particular las de altas implicancias, tienen llegada anual con sus reportes a todas las escuelas del sistema, y en algunos casos a todos los alumnos de los grados evaluados (especialmente en los sistemas que aplican pruebas de egreso de la educación media). Los países que realizan operativos muestrales no suelen enviar los datos obtenidos a las escuelas de la muestra: en su lugar, analizan los datos y producen y distribuyen publicaciones de interés para diferentes audiencias pedagógicas y políticas.

En cualquiera de los dos casos, con censos o con muestras, la entrega de reportes y publicaciones por sí sola no es suficiente para que la información se comprenda y sea utilizada. A menudo la entrega se produce en el contexto de jornadas de capacitación docente, reuniones de supervisores, conferencias de prensa, o "tours" de los equipos técnicos y sus asistentes por diferentes regiones del país para comentar y analizar los resultados. Estos esfuerzos se vuelven cada vez más necesarios conforme se complejizan los modelos de evaluación y los tipos de datos que ellos ofrecen. No hace falta mucha explicación, con algunas audiencias, para comprender que una provincia o departamento logró en promedio el 30% de respuestas correctas en matemática. Pero para comprender cómo diferentes subpoblaciones de estudiantes de una provincia alcanzan diferentes niveles de logro, referidos estos a competencias disciplinares de alta densidad conceptual y procedimental, en un recorrido curricular que no distingue necesariamente grados de escolaridad, sino tramos en un continuo de saberes progresivos y de creciente complejidad, algunas explicaciones más personalizadas son necesarias.

La difusión de los datos se ve facilitada, al menos en teoría, a medida que los programas de evaluación van utilizando recursos informáticos y de Internet para estos fines. En casi todos los casos nacionales, las unidades de evaluación tienen sitios web propios o alojados en algún portal oficial, desde donde se puede consultar los resultados de diferentes subsistemas, resultados de escuelas, contenidos evaluados e incluso los ítems o reactivos utilizados durante las pruebas. Dado que en la mayor parte de los países se quiere evitar el efecto *ranking* o de posible estigmatización, los resultados de alumnos o instituciones participantes en operativos censales solamente son accesibles mediante el uso de una contraseña.

En general, estas informaciones se ofrecen para ayudar a comprender mejor los resultados y los referentes curriculares de la evaluación. También persiguen ese propósito los países que, luego de los operativos oficiales, "liberan" algunos instrumentos o colección de ítems para que las escuelas puedan aplicarlos de forma autónoma y así familiarizarse más directamente con los modelos, los contenidos y los criterios de corrección que se utilizan en mediciones a gran escala.

PRINCIPALES AVANCES EN LA EVALUACIÓN Y RETOS DE FUTURO

La descripción de los sistemas y programas nacionales de evaluación presentada hasta este punto no es, ni pretende ser, exhaustiva. La diversidad de enfoques de evaluación, de metodologías, de formas de reporte y entrega de resultados, así como de los contextos educativos y políticos en que se presentan, es tan inmensa como imposible su descripción en este texto. Aun así, y arriesgando la sobre-simplificación del tema, los avances más importantes en materia de evaluación estandarizada en la región durante las dos últimas décadas podrían resumirse de la siguiente manera:

- *Fortalecimiento institucional.* Los sistemas nacionales de evaluación han cobrado mayor estabilidad, continuidad y organicidad de acciones dentro de las administraciones del sector. Los equipos técnicos trascienden los cambios de gobiernos, acumulando experiencia y desarrollando *expertise* que los hace menos vulnerables a los recambios de autoridades y personal técnico de los ministerios u otras agencias oficiales.
- *Integralidad de propósitos y acciones.* Conforme ganan institucionalidad y experiencia, los programas de evaluación van abarcando otras áreas del quehacer educativo y no solamente la medición del logro académico. El desarrollo técnico y la capacidad de conceptualizar la complejidad del sistema educativo va permitiendo estudios cada vez más integrales de los factores y procesos que influyen en el aprendizaje y en la calidad educativa en general.
- *Experiencia internacional.* La vinculación con organismos internacionales y la participación en programas de evaluación internacional comparada han ofrecido importantes oportunidades de capacitación, desarrollo técnico y metodológico, así como acceso a nuevos modelos de evaluación y formas de optimizar los referentes del currículo.
- *Modelos, enfoques e instrumentos.* La actualización de los propósitos de la evaluación, enfocada cada vez más en la provisión de informaciones útiles para el mejoramiento del sistema educativo y de la gestión pedagógica en particular, ha movilizado el desarrollo de nuevos enfoques e instrumentos de medición. Destacan principalmente la transición a pruebas referidas a criterios, la determinación de niveles de desempeño y la medición del valor agregado que aportan las escuelas de diferentes contextos. Esto último implica un mejor uso de los datos escolares y extraescolares que se recogen mediante cuestionarios aplicados simultáneamente con las pruebas de aprendizaje, así como la necesidad de realizar análisis longitudinales con instrumentos validados para ese fin.
- *Formas de reporte.* El uso de nuevos modelos e instrumentos de evaluación, que a su vez generan datos de creciente complejidad y sustancia, va impulsando el diseño de reportes de resultados más completos, más comprensibles y potencialmente más útiles para una amplia variedad de audiencias.
- *Difusión y uso de los resultados.* Se observa una creciente preocupación y acciones más integrales para hacer de la entrega de los resultados una instancia *formativa* y no meramente informa-

tiva. Se van diversificando los canales de entrega, tanto de forma presencial como mediatizada por Internet y otros recursos digitales e informáticos.

Como puede verse, los logros alcanzados en materia de evaluación estandarizada en la región no son menores, especialmente si se considera que el mayor desarrollo ha ocurrido durante un lapso de 15 a 20 años. La tarea, sin embargo, podría decirse que acaba de comenzar.

Son múltiples los desafíos que enfrentan los sistemas nacionales de evaluación para los próximos años. Aunque en muchos casos los equipos y sus metodologías se han afianzado, y existen ya ciertas garantías de continuidad en su labor, los nuevos escenarios socioeducativos, las demandas políticas y la propia ética profesional seguirán exigiendo cada vez mejores evaluaciones y mejores estrategias para hacer un uso responsable y efectivo de los datos e informaciones producidos.

Será necesario seguir revisando, actualizando y socializando los referentes curriculares sobre los que se evalúan los aprendizajes, de manera tal que los resultados cobren sentido pedagógico y despierten iniciativa de mejora en una mayor cantidad de actores. Si los currículos nacionales siguen considerándose como amplios lineamientos para la enseñanza que cada escuela y cada docente recortan y ajustan según su criterio, difícilmente se lograrán acuerdos acerca de qué tipos de aprendizajes deben ser alcanzados por todos los alumnos del sistema. La evaluación a gran escala puede aportar mucho en esta dirección, en la medida en que sus instrumentos prioricen o jerarquicen algunos contenidos y logren transmitir un mensaje claro y unívoco respecto a qué tipos de saberes todos los alumnos tienen el derecho a acceder. Para esto, el trabajo conjunto entre evaluación y los ámbitos de decisión en materia de currículo es imprescindible.

La adopción de referentes curriculares más complejos y de mayor especificidad tendrá beneficios para los programas de evaluación, toda vez que un referente claro y operacional ayuda en la delimitación y descripción de los contenidos que serán objeto de medición. Pero también impondrá desafíos importantes en la construcción de reactivos e ítems de calidad que, individualmente y en conjunto, puedan captar con validez la evidencia de aprendizajes complejos.

También impone un desafío importante la medición de los factores asociados al rendimiento. Aunque muchos países, si no todos, aplican cuestionarios de contexto, frecuentemente los datos recogidos se desaprovechan o resultan directamente inútiles para cruzarlos estadísticamente con el rendimiento académico. Un trabajo idóneo de análisis de factores asociados, que permita medir con validez el progreso de los aprendizajes en el tiempo, o que habilite el reporte de valor agregado, requiere un modelo cuidadosamente diseñado y el uso de instrumentos comparables o equivalentes entre aplicación y aplicación. Para mejorar todos estos aspectos técnicos, seguirá siendo necesaria la capacitación profesional continua de los equipos nacionales, así como la adopción y sujeción a los estándares técnico-metodológicos en medición que ofrece actualmente la literatura internacional especializada.

A las altas autoridades políticas y técnicas del sector les deparan responsabilidades tanto o más importantes que las anteriores. En primer lugar, deberán evitar demandas de evaluación urgentes que pongan en riesgo la calidad de los operativos, que desgasten inútilmente los recursos humanos y que impliquen inversiones de dinero no justificadas. Por el contrario, les cabe la responsabilidad de familiarizarse primero con los modelos e instrumentos existentes, y en todo caso proponer mejoras y movilizar cambios que guarden coherencia con las políticas que se busca implementar y evaluar, siempre que exista la capacidad técnica para hacerlo o los recursos para crearla.

Valga, para concluir, una reflexión sobre el papel que juegan los docentes en este escenario de evaluación estandarizada en la región. Si consideramos, como repetidamente se afirma, que las reformas y el cambio educativo solo son exitosos si se materializan en el aula, los avances en materia de evaluación no pueden ser la excepción. Por mejores modelos y datos que obtengamos, por mejores y racionales decisiones de política que se tomen sobre los resultados, poco cambiará si los docentes no logran, dentro de su compleja labor didáctica, realizar una mejor evaluación de los aprendizajes.

En la actualidad, maestros y directores de escuela, supervisores, e incluso los formadores de docentes, carecen de los conocimientos conceptuales y metodológicos básicos para realizar evaluaciones objetivas, criteriales y coherentes con las tendencias pedagógicas y curriculares vigentes. Quienes tenemos la oportunidad de hacer docencia y capacitación en evaluación, nos encontramos sembrando prácticamente en terreno virgen. La gran mayoría de docentes y pedagogos nunca han sido puestos en situación de diseñar un referente curricular sólido para la evaluación y rara vez están familiarizados con el concepto de "constructo". En su capítulo, Juan Manuel Esquivel trata este tema de forma más detallada. Solo mediante prácticas intensas logran operacionalizar una competencia compleja en unidades conceptuales y procedimentales que puedan medirse con instrumentos de aula. Con dificultad, y tras analizar un vasto número de ejemplos, pueden construir indicadores alineados con el constructo en juego. Luego queda la tarea de diseñar preguntas, actividades y situaciones de evaluación variadas que puedan captar pruebas de logro o aprendizaje. Aun así, tras semanas de trabajo, ensayo y error, reconocen que apenas han vislumbrado una nueva forma de evaluar y de pensar el currículo y la enseñanza.

Esto nos hace reflexionar sobre el potencial de los sistemas nacionales de evaluación para contribuir a una mejor educación y nos obliga a preguntarnos por qué todavía la evaluación estandarizada, aun cuando sea idónea y aparezca cargada de datos pedagógicos interesantes, resulta tan ajena en el ámbito local y escolar. Evidentemente, ese espacio que los programas de evaluación se han ganado en el sistema educativo todavía puede y debe extender sus fronteras. Y a quienes argumentamos que ese potencial y esa posibilidad existen, nos cabe la responsabilidad compartida de tender los puentes necesarios para que suceda.

Anexo

Participación de América Latina, España y Portugal en pruebas internacionales

PISA (OCDE)

2000: Argentina, Brasil, Chile, México, Perú, [España, Portugal].

2003: Brasil, México, Uruguay, [España, Portugal].

2006: Argentina, Brasil, Chile, Colombia, México, Uruguay, [España, Portugal].

2009: Argentina Brasil, Chile, Colombia, República Dominicana, México, Panamá, Perú, Uruguay, [España, Portugal].

TIMSS (IEA)

1995: Colombia, México, [España, Portugal].

1999: Chile.

2003: Chile.

2007: Colombia, El Salvador.

PIRLS (IEA)

2001: Argentina, Colombia.

2006: [España].

Educación Cívica (IEA)

1999-2000: Colombia, Chile, [Portugal].

Estudios del LLECE (UNESCO-OREALC)

1997 (Primer Estudio Internacional Comparativo): Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica*, Cuba, República Dominicana, Honduras, México*, Perú*, Venezuela [países marcados con * participaron, pero no divulgaron resultados].

2007 (SERCE-Segundo Estudio Regional Comparativo y Explicativo): Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, República Dominicana, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú y Uruguay. Estado mexicano de Nuevo León.

Las múltiples caras de la evaluación

Indicadores y desafíos de los sistemas educativos en América Latina

Margarita Poggi

INTRODUCCIÓN

El presente capítulo aborda el tema de los indicadores y su relación con los principales desafíos que enfrentan los sistemas educativos en América Latina en la actualidad. Partiendo de la importancia de producir información sobre estos sistemas, una cuestión clave está vinculada con las preguntas relevantes sobre las que deben informar los indicadores educativos. Por ello se analizan algunos de los principales desafíos que enfrentan en la actualidad los sistemas educativos de América Latina, ya sea que remitan a problemas históricos o a situaciones emergentes de las nuevas dinámicas sociales y educativas. Ellos se vinculan con preguntas relevantes sobre las que debe disponerse de información tanto para el planeamiento de las políticas como para su evaluación. Por último, el capítulo concluye con una mirada hacia los desafíos que se plantean en la producción, el análisis y el uso de la información educativa en relación con los temas anteriormente mencionados y a partir de una visión general sobre los avances, pero también los desafíos pendientes, que tienen las oficinas productoras de conocimiento sobre la marcha de los sistemas educativos en los países de la región.

LA INFORMACIÓN SOBRE LOS SISTEMAS EDUCATIVOS

Distintos actores sociales (tomadores de decisión en política educativa, equipos técnicos que colaboran en Ministerios de Educación, docentes y alumnos y la comunidad en términos amplios) tienen cada vez más claridad respecto a la complejidad de los sistemas educativos. Distintos factores permiten explicar las razones de esta afirmación: por un lado, tiene lugar un proceso de transformación de sujetos con culturas de origen social y familiar diversas en la medida en que se transmiten contenidos relevantes y se forman capacidades; por ello, la temporalidad generalmente remite al largo plazo, tanto para producir cambios en su estructura y dinámica interna, como para el proceso de formación del conjunto de sujetos que se escolariza a lo largo de la vida. Por otra parte, la dimensión y escala se han multiplicado a la par del crecimiento de las demandas por la democratización, esto es, la ampliación de la cobertura. Por último, existe una multiplicidad de organizaciones y actores que están implicados en el proceso educativo, con objetivos diversos, en ocasiones coincidentes, en otros casos, contradictorios entre sí, razón por la cual la variedad y cualidad de las demandas no solo se han ampliado, sino que resultan cada vez más diversificadas.

Además, la organización y dinámica de los sistemas educativos y de sus niveles no siempre ha respondido a una previa planificación. Por el contrario, en general las escuelas que integran los diversos niveles (y estos mismos) preexisten a la organización de los sistemas como tales.

Así, los tramos organizados en niveles y las modalidades u orientaciones que algunos de estos presentan se han desarrollado históricamente con una tradición propia, generando subculturas académicas diferenciadas para cada uno de ellos. Esto explica asimismo, en parte, la existencia de normativas y regulaciones no siempre articuladas para los distintos niveles.

No escapa a este sistema complejo la dificultad que plantea la elaboración de políticas públicas para su gobierno y administración, y la relevancia de la información para su diseño, seguimiento, monitoreo y evaluación.

Así, en las últimas décadas en los Ministerios de Educación de los países de América Latina se ha promovido la generación o fortalecimiento de sistemas de información, que incluyen tanto las llamadas estadísticas educativas como la evaluación de la calidad. Las principales funciones que se les reconocen a estos sistemas son las siguientes: a) permiten detectar y volver visibles las principales tendencias y logros de los sistemas educativos; b) contribuyen a objetivar la discusión sobre determinadas temáticas; c) permiten profundizar la comprensión de los problemas de los sistemas, subsistemas e instituciones; d) con ello, orientan la selección de prioridades en el diseño de políticas educativas, y e) permiten comunicar de forma sintética aspectos o parcelas de un sistema complejo.

Los indicadores resultan cada vez más necesarios para construir miradas analíticas sobre un sistema educativo y por ello forman parte de un sistema de información. Resulta importante distinguir entre un sistema de información estadística, el cual incluye todos los datos educativos en una base única y extensiva, y un sistema de indicadores, que tiene por objetivo central considerar y analizar tendencias en torno a ciertas cuestiones de un sistema educativo y revelar problemas (Sauvageot, 1997). En ese sentido, un indicador es "un artificio que proporciona información relevante acerca de algún aspecto significativo de la realidad educativa" (Tiana, 1997) y como tal resulta siempre de una construcción y un consenso. Al mismo tiempo, estos y otros autores señalan que es importante el diseño de *sistemas* de indicadores, y no una colección de ellos, para generar información más amplia y precisa sobre un sistema educativo. Mayor es la discusión, sin embargo, sobre la cantidad pertinente de indicadores para dar cuenta de sus principales tendencias.

En América Latina pueden hallarse, además de los sistemas nacionales de indicadores, varias iniciativas orientadas a la producción y análisis de información. Sin pretensión de ser exhaustivos, pueden mencionarse los siguientes. El sistema de indicadores de la UNESCO para el seguimiento y monitoreo del cumplimiento de las metas de Educación para Todos y los Objetivos de Desarrollo del Milenio, en el marco de su Encuesta Anual de Educación y el Proyecto Mundial de Indicadores (WEI). Más específicamente, los desarrollos del Sistema Regional de Información de la OREALC/UNESCO Santiago para el seguimiento en la región de los compromisos en materia educativa, tanto en el marco de EPT/PRELAC como de las metas de las Cumbres de las Américas, este último proyecto (PRIE) bajo la coordinación general de la SEP-México. Por otra parte, el Sistema de Información de Tendencias Educativas en América Latina (SITEAL) es una iniciativa conjunta de la OEI y el IIPE/UNESCO Buenos Aires y se centra en la relación entre los procesos educativos y los contextos sociales y culturales y el análisis de las desigualdades sociales y educativas. En cuanto a los estudios internacionales y regionales de evaluación, hay diversos países de la región que participan en ellos y que permiten también consolidar información sobre logros de los estudiantes y las variables contextuales y escolares asociadas a ellos. En el capítulo de Enrique Roca de este libro se analizan estos estudios con más detalle.

A continuación se analizarán los principales desafíos que enfrenta la educación en América Latina, los cuales remiten tanto a algunos problemas recurrentes como a otros que son más recientes, en la medida en que la dinámica de los sistemas educativos se encuentra atravesada por las demandas sociales que deben enfrentar. Se parte de estos desafíos, porque ellos generan nuevos requerimientos a los sistemas de información, que serán abordados más adelante.

PRINCIPALES TENDENCIAS Y DESAFÍOS DE LA EDUCACIÓN EN AMÉRICA LATINA[1]

Puede observarse que los grandes temas de las políticas educativas de la región remiten en un cierto sentido a viejos problemas de sus sociedades y sistemas educativos. No obstante, su persistencia en las agendas se vincula con dos cuestiones. La primera tiene que ver con los intentos fallidos (algunos parcialmente, es cierto) para resolverlos. Pero, por otro lado, también es necesario destacar que, sobre un telón de fondo semejante, esos problemas van adoptando nuevas manifestaciones o formas. Su presencia en las agendas requiere actualizar los diagnósticos y hacerlos más pertinentes a la situación actual, lo que sin duda significa una mirada mucho más precisa para comprender el contexto y también el diseño de estrategias más consistentes con esta mirada.

Asimismo, es importante señalar que no se agotarán todos los temas que podrían integrarse a una agenda ni se atenderán las problemáticas educativas específicas de cada uno de los países de la región. Por el contrario, se pretenden señalar las tendencias prioritarias para las agendas de política educativa que plantean nuevas preguntas a los sistemas de información. Los indicadores educativos deben aportar información para responder estas preguntas, asegurando entonces tanto la posibilidad de efectuar análisis en estudios longitudinales (que permitan analizar la evolución en el tiempo) como sincrónicos (que posibiliten la comparación transversal entre sistemas o subsistemas educativos).

En términos generales, es importante destacar no solo el alto grado de heterogeneidad entre los sistemas educativos de los países de la región, sino fundamentalmente el alto grado de heterogeneidad dentro de cada país cuando se consideran distintos contextos (urbano/rural, grandes urbes, acceso y egreso diferenciado por sector social, pertenencia a grupos originarios, etc.).

En cuanto a las principales tendencias generales de la situación educativa de América Latina, pueden señalarse las cuestiones que se mencionan a continuación.

La desigualdad social y educativa

En los países de la región la desigualdad social se encuentra estrechamente relacionada con las posibilidades y los límites que circunscriben los márgenes de actuación de las políticas educativas. Hay ya acumuladas pruebas suficientes para demostrar que la probabilidad de inserción en la educación escolar es más elevada en el quintil más rico de la población (en especial cuando se la compara con la que tiene el quintil más pobre). En efecto, salvo en el nivel primario, que está prácticamente universalizado, en el nivel inicial y en el secundario las tasas de cobertura están directamente relacionadas con el nivel de ingreso de las familias y con el capital escolar de la familia del niño o joven. Asimismo, es preciso tener en cuenta que la probabilidad de acceder,

[1] Este apartado retoma aspectos desarrollados en Poggi, M., *Old and new perspectives for traditional issues in planning*, París, IIEP-UNESCO (en prensa).

permanecer y aprender en la escuela es significativamente más alta en la población urbana que en la población rural. Por otra parte, las tendencias muestran que las desigualdades educativas tienden a desplazarse hacia los niveles superiores del sistema educativo (secundario superior y universitario) y hacia la etapa de ingreso a este (tres y cuatro años de edad).

Es necesario señalar que, en el contexto de sociedades extremadamente desiguales, el aumento de la escolarización en el nivel medio se acompaña de una fuerte acentuación del carácter estratificado de la oferta institucional de enseñanza. El carácter territorial de la oferta educativa refuerza esta tendencia a la segmentación escolar. En ciertos contextos, la fragmentación y jerarquización de los sistemas escolares ha alcanzado tal magnitud y calidad que resulta cada vez más difícil pensarlo como un "sistema" homogéneo que tiende a cumplir las mismas funciones y a alcanzar los mismos objetivos. La desigualdad de la escolarización en cada nivel del sistema educativo determina fuertemente la probabilidad de acceso y terminación de estudios en los niveles subsiguientes, principalmente por los conocimientos que transmite y las habilidades cognitivas y actitudinales que forma en los estudiantes.

Entre las estrategias priorizadas en la región, puede encontrarse que una de ellas ha sido diversificar y flexibilizar la oferta educativa para atender a la diversidad. No obstante, cabe llamar la atención porque esa diversificación muchas veces deviene una oferta desvalorizada que, más allá de sus loables intenciones, profundiza las diferencias educativas.

Aun cuando se ha señalado en numerosas ocasiones que la acción efectiva que permite intervenir para reducir las diferencias sociales y educativas requiere políticas integrales e intersectoriales, es necesario señalar los enormes esfuerzos que estas requieren y las dificultades que presentan para su implementación. No obstante, del análisis de las políticas públicas que se desarrollan en países de la región se infiere que sigue constituyendo una aspiración pertinente, que no anula ni deslegitima la capacidad de los Ministerios de Educación para intervenir en la disminución de estas desigualdades desde los ámbitos de su incumbencia. De esta manera, la democratización no se entenderá solo como una prolongación de las trayectorias escolares, sin que se atiendan de igual modo otros de sus aspectos fundamentales: esto es asegurar, de modo efectivo la igualdad de oportunidades en el acceso al conocimiento y la igualdad de logros, aspectos que constituyen el horizonte de las políticas educativas y están en estrecha relación con el tema de la calidad.

La calidad educativa

Existe un consenso en criticar la concepción de calidad difundida en la región en los ochenta y, fundamentalmente, en los noventa, por considerarla simplificadora y reductora de las múltiples dimensiones a las que hace referencia. Diversos países en la región vienen realizando esfuerzos por trabajar en esta línea. Puede, en este sentido, señalarse el caso de México, en el cual esta preocupación se ha planteado hace ya varias décadas. Asimismo, OREALC/UNESCO Santiago ha construido una posición que se plasma en la concepción de calidad educativa propuesta en la última reunión del PRELAC (2007), en donde se expresa que "la calidad de la educación en tanto derecho fundamental, además de ser eficaz y eficiente, debe respetar los derechos de todas las personas, ser relevante, pertinente y equitativa. Ejercer el derecho a la educación es esencial para desarrollar la personalidad e implementar los otros derechos".

Esta concepción supone una mirada más compleja hacia el tema de la calidad; subraya el derecho a la educación, el cual está fundado en los principios de obligatoriedad, gratuidad y en el derecho

a la no discriminación. Otra dimensión es la relevancia, que remite principalmente al desarrollo de saberes y competencias necesarias para participar en los diferentes ámbitos de la vida social y construir proyectos con relación a los otros. La relevancia también puede analizarse en conexión con las finalidades educativas que, propias de un momento y contexto dados, se enmarcan en un proyecto político y social y dan sentido a las prácticas que se desarrollan en las escuelas que integran un sistema educativo. Por otra parte, pertinencia y equidad suponen atender, en el primero de los casos, a la diversidad de necesidades de los individuos y de los contextos, para que la educación sea significativa para personas provenientes de distintos estratos sociales y culturas; en el segundo, significa asegurar la igualdad de oportunidades para acceder a una educación de calidad para toda la población, asegurando las condiciones (recursos y ayudas) que requieren.

La preocupación por aspectos más amplios que los tradicionalmente considerados en la concepción de la calidad educativa (eficacia y eficiencia), como son los mencionados, supone en consecuencia un interesante replanteo de base que tiene efectos en la definición de las políticas para la región. En este sentido puede afirmarse que, aun cuando el problema de la calidad educativa ha estado en las agendas en las últimas dos décadas, ha adoptado en los años recientes una complejidad que exige no solo estrategias diferentes para su evaluación (y con ello una actualización de los sistemas nacionales de evaluación), sino fundamentalmente abordajes mucho más integrales para su mejora.

La calidad del aprendizaje es siempre el resultado de una combinatoria de factores escolares y sociales relacionados entre sí. Además de las condiciones sociales y familiares, hay condiciones de aprendizaje que son fundamentales, vinculadas con el interés por aprender que se construye también como resultado de buenas prácticas escolares. A ello se suman factores sistémicos (recursos edilicios y equipamiento escolar, condiciones laborales docentes, diseños curriculares actualizados y pertinentes, estrategias de seguimiento de los procesos escolares de los niños y jóvenes con dificultades para aprender, etc.) como una buena escuela, docentes con una buena formación, estrategias y métodos didácticos adecuados a las características de la población de estudiantes.

En la región, solo en los últimos años, el tema de la calidad educativa desde esta visión más amplia y compleja se fortalece en las agendas de los ministerios. En parte, porque la experiencia acumulada ha marcado los límites de reducirla solo a los resultados académicos (aunque obviamente no pueden quedar excluidos); en parte también porque, aun cuando estos resultan una dimensión ineludible del tema, los factores sociales y educativos que intervienen en ella, como se ha dicho, requieren políticas más integrales que las que se promovieron.

El financiamiento educativo

En promedio, América Latina asigna aproximadamente un 4,7% de su PIB a educación en la actualidad. Como región, se encuentra en segundo lugar por detras de América del Norte y Europa Occidental, que destinan un 5,7% anual en promedio. Si bien hay datos empíricos en la región de que la inversión sectorial global creció en los noventa, en muchos casos no hizo más que compensar lo que se había perdido en la década anterior (Morduchowicz y Duro, 2007). No obstante, para analizar el significado que estos porcentajes suponen, debe tenerse en cuenta que estos son solo una de las medidas del esfuerzo financiero de cada país en sus sistemas educativos y que no deben analizarse fuera de su contexto. En algunos, pueden resultar elevados por el tamaño de sus economías o por su menor población escolar. En otros, pueden resultar escasos, cuando se analizan la

complejidad y envergadura de los problemas que deben atender. Por ello, sigue constatándose que los esfuerzos financieros son dispares y el promedio es bajo, especialmente cuando es comparado con la inversión de los países de mayor desarrollo.

Es importante también superar los planteos reduccionistas que limitan la discusión solo al uso más eficiente de los recursos; especialmente porque ha dado lugar a distintas controversias. En el mismo sentido, más allá de que deba reducirse el nivel de ineficacia en la utilización de los recursos disponibles, a la vez es necesario aumentar de forma significativa la inversión para el sector, especialmente si se quieren cumplir algunas de las metas fijadas en distintos acuerdos internacionales. Los países en la región ya están desarrollando algunas estrategias, entre ellas nuevas leyes de financiamiento o de asignación diferente de los recursos del sector, en donde Argentina, Brasil, Colombia y Chile son claros ejemplos de ello. Estas normativas requieren una profundización de las tareas de planeamiento para hacer efectivas y viables las metas que ellas plantean.

Si bien se han mencionado hasta aquí tres grandes tendencias que caracterizan los principales desafíos de los sistemas educativos de la región, pueden destacarse también algunos temas más específicos, los cuales adoptan manifestaciones distintas según cada sistema y la oferta que caracteriza cada país.

El reconocimiento de "nuevos" sujetos en las instituciones y los sistemas educativos

Se está asistiendo en los sistemas educativos de la región a la necesidad de estrategias de escolarización para atender a la diversidad que caracteriza a la nueva población que se ha integrado a los sistemas. En sentido estricto, no se trata de "nuevos" sujetos, ya que hace décadas que se encuentran asistiendo a las escuelas, pero sí puede afirmarse que hasta cierto punto han permanecido invisibles en ellas. Algunos ejemplos, muy distintos entre sí, permiten dar cuenta de esta cuestión: los alumnos y estudiantes con marcada sobre-edad en primaria y secundaria, provenientes de los sectores más pobres (sobre-edad que se produce como un efecto combinado de la repitencia recurrente o abandonos temporarios y reingresos al sistema); la presencia creciente de jóvenes en la educación de adultos, en parte por cuestiones sociales, pero también por las condiciones de la escolarización en el nivel secundario, que ha permanecido casi inmutable a pesar de los intentos por transformarlo; alumnos que pertenecen a distintos pueblos originarios o provenientes de otros países de la región con características culturales propias. Todos estos grupos requieren la puesta en marcha de estrategias institucionales y pedagógicas que permitan atender a su especificidad para asegurar su proceso de escolarización.

La primera infancia y su incorporación más temprana en los procesos educativos

En general, puede constatarse en la región que, en las políticas de los gobiernos, se está otorgando mayor prioridad a la atención y educación de la primaria infancia, en contraste con la situación pasada, en donde claramente el tema tenía menor importancia que otros objetivos educativos. Cabe señalar que dentro de los países se dan grandes diferencias en cuanto a la oferta: son los niños de las familias más pobres y de las que viven en zonas rurales los que tienen muchas menos posibilidades de acceso, cuando en realidad se trata de quienes más beneficios podrían obtener de los programas de atención educativa temprana.

La educación de la primera infancia recibe en este contexto una atención particular en las políticas del sector, caracterizada por una concepción integral. En este sentido, involucra no solo al sistema educativo formal (particularmente el nivel inicial), sino que se concibe como un desafío que excede a la política educativa porque exige articulaciones con los sectores de salud, atención a las familias, desarrollo comunitario, etc. Constituye el ejemplo por excelencia en el cual sin acciones integrales, conjuntas y coordinadas de distintos sectores de gobierno no se puede garantizar una política de equidad. En lo que respecta al sistema educativo en particular, la ampliación de la oferta de la educación inicial, en algunos casos, y de la obligatoriedad en la asistencia, en otros, refleja la importancia del tema para la región.

La ampliación en el acceso y la mejora del egreso en educación secundaria con aprendizajes relevantes

Cuando se analiza el patrón de la expansión de la escolarización en el nivel medio, especialmente en la década de los noventa, es posible afirmar que se ha registrado en un tiempo relativamente corto y, como consecuencia de ello, los sistemas educativos no han estado siempre en condiciones de dar las respuestas más adecuadas a esta demanda. Por un lado, puede constatarse una tendencia en la región a universalizar y, en algunos casos, hacer obligatoria, la educación secundaria (como en Argentina y Chile). Por el otro, se ha detenido el crecimiento en la ampliación de la cobertura, característico de la década pasada, lo que permite fundamentar una hipótesis vinculada con los límites a la expansión del nivel. Obviamente, esta puede explicarse en parte por factores exógenos a los sistemas educativos, pero no debe descartarse la incidencia de las condiciones propias de la oferta existente para escolarizar jóvenes provenientes de sectores muy heterogéneos en términos de sus características socioeconómicas. Además, las culturas juveniles propias de esta etapa vital configuran características muy particulares en el momento de la asistencia al nivel secundario. Es este también otro ejemplo de los "nuevos" sujetos presentes hoy en las instituciones educativas, que se mencionó anteriormente.

En relación con lo expresado, puede anticiparse en consecuencia que los próximos pasos para la escolarización en la educación media (especialmente en la educación secundaria superior) requerirán esfuerzos de mucha mayor envergadura tanto en materia de expansión de la oferta (infraestructura, materiales didácticos, formación de docentes, etc.) como en la adecuación de las estrategias y los modelos pedagógicos. La educación secundaria deberá innovar tanto en cuanto a su formato organizacional, como en las estrategias y procedimientos de enseñanza y de aprendizaje, para asegurar el desarrollo de conocimientos y actitudes que habilitan para el trabajo y la ciudadanía activa. Por ello, en este contexto, la cuestión pedagógica adquiere una importancia estratégica fundamental y el desarrollo de propuestas innovadoras y fundamentadas en experiencias resultará importante para dotar con nuevas respuestas a la tradicional escuela secundaria.

La educación de jóvenes y adultos

Tradicionalmente, la educación de adultos se circunscribía, en primer lugar, a asegurar la alfabetización que permitía, a aquellos que no habían tenido la oportunidad de asistir a la educación básica, acceder a los conocimientos fundamentales que supone el dominio de la lengua escrita y del cálculo. Por ello se fueron organizando ofertas específicas que se hicieron cargo de la alfabetización inicial y luego de la educación básica o primaria de la población adulta. En la medida en que la oferta educativa se expandió en las últimas décadas, se fortaleció también la de la educación se-

cundaria para permitir la continuidad de los estudios de los adultos que concluían la primaria, o que habían abandonado la secundaria, y podrían así reinsertarse en el sistema educativo formal.

No obstante lo señalado (que se caracteriza a su vez por matices importantes entre los países de la región, vinculados con las importantes distancias en cuanto a la situación de este subsistema), se presentan fuertes pruebas de que la educación secundaria para adultos incorpora cada vez más a jóvenes que ingresaron a la secundaria común, pero no pudieron permanecer en ella y aún menos egresar del nivel. Se puede anticipar incluso que esta tendencia irá en aumento en los próximos años, cuando se analizan los grupos de jóvenes que no tienen estudios secundarios completos. Por ello es esperable que en los próximos años se incremente considerablemente la demanda de este tipo de oferta educativa, de modo que permita a esos jóvenes acceder a los conocimientos y la certificación de los estudios secundarios. Ello plantea un desafío importante en las agendas educativas, ya que requerirá el diseño de programas y ofertas que consideren las nuevas manifestaciones de esta problemática en la oferta para jóvenes y adultos.

En fin, no se detallarán aquí otros temas específicos que pueden detectarse cuando se analizan diferentes temáticas educativas, entre los que pueden mencionarse los siguientes: los docentes, su formación y desarrollo profesional; la educación superior; la dinámica entre la educación y las tendencias de desarrollo social y económico; el gobierno y la administración de los sistemas educativos. Cada uno de ellos requeriría un tratamiento que excede los límites del presente capítulo. Solo se han destacado algunas temáticas para dar cuenta de ciertas problemáticas y sus manifestaciones específicas en la actualidad y porque ellas además están planteando nuevos requerimientos a los sistemas de información.

PRINCIPALES RETOS DE LOS SISTEMAS DE INDICADORES DE LA REGIÓN

Ya ha sido señalado anteriormente que un sistema de indicadores debe responder a preguntas relevantes que distintos actores se formulan con respecto a la organización y evolución de un sistema educativo (y de sus subsistemas). Por ello, en el marco de un nuevo escenario regional social y educativo que viene configurándose en los últimos años y, al mismo tiempo, sobre la experiencia de las últimas décadas con respecto al desarrollo de los sistemas de indicadores, pueden analizarse los desafíos que estos tienen de cara al futuro en relación con tres aspectos clave: aquellos vinculados con a) la recolección y producción; b) el análisis, y c) la difusión y el uso de la información social y educativa.

Corresponde realizar una primera aclaración: es cierto que solo con fines analíticos pueden separarse tres etapas en las cuales se toman decisiones con respecto a la información sobre un sistema educativo, pero tienen implicancias directas en el conjunto del proceso. No obstante, para los fines de esta presentación, se separarán artificialmente para su mejor organización. Cabe aclarar asimismo que algunos desafíos que se mencionan en los siguientes puntos son más estructurales[2], mientras que otros se encuentran más estrechamente relacionados con las nuevas preguntas y, en consecuencia, los nuevos problemas que resultan de la dinámica propia de los sistemas educativos de la región, los cuales se abordaron en el apartado anterior.

[2] Puede ampliarse este apartado en el informe del PRIE (2007), *Balance de las misiones de diagnóstico de los sistemas de información educativa*, resumen ejecutivo, UNESCO y Secretaría de Educación Pública de México. Ver en: http://www.unesco.org/santiago.

Sobre la recolección y producción de información

Mejorar las metodologías para la recolección de información.

Este aspecto es clave en la definición del marco conceptual y metodológico de un sistema de indicadores, pero también lo es a medida que este se desarrolla y se releva información de modo continuo y sistemático. Además está vinculado con la calidad de la información elaborada, asegurando la confiabilidad y representatividad de los datos.

Un problema recurrente que se encuentra en este aspecto es la "confrontación" entre los datos "duros" en relación con los datos "impresionistas", propios de actores del sistema educativo. La distinción entre objetos empíricos y objetos teóricos (INEE, 2007) ayuda a resolver esta confrontación ya que, cuando hablamos de información estadística, se trata de construir objetos teóricos, orientados por ciertos propósitos, diferenciándolos de los innumerables objetos empíricos que se encuentran en la realidad.

Por otra parte, el problema de la verificación resulta siempre clave, lo que implica anticipar los procedimientos de supervisión y verificación en la etapa de recolección de la información. La sistematización y documentación del proceso de construcción de un sistema de indicadores solo recientemente se incorpora como un ejercicio sistemático en las oficinas productoras de información y tiene implicancias tanto a la hora de recolectar como de interpretar la información en los distintos niveles de agregación que se hayan previsto.

Prestar especial atención a los niveles de desagregación de la información disponible y dar respuesta a requerimientos de relevamientos específicos de información que atiendan a las necesidades nacionales y subnacionales (provincia/estado, municipio, según organización de un sistema educativo).

Si bien un sistema de información da cuenta de las tendencias generales de un sistema educativo (o de algunos de sus subsistemas, niveles, modalidades), debe tenerse en cuenta también el alto grado de heterogeneidad propia de cada uno de los países de la región, al que se hizo referencia anteriormente. Por ello, este aspecto (los niveles de desagregación de la información) se vuelve especialmente relevante para orientar procesos de decisión de política educativa.

Por otra parte, pueden surgir necesidades de producción de nuevos indicadores, cuestión que pone en discusión el tema del número adecuado de indicadores que debe integrar un sistema. Si bien a nivel de un sistema nacional de indicadores la discusión parece saldada, ya que se cuenta con relevantes experiencias en la producción de estos, cuando se desagrega la información disponible para dar cuenta de la heterogeneidad interna de un sistema educativo, es posible que se presenten nuevos problemas para resolver en la gestión y, con ello, la necesidad de producir nueva información y generar nuevos indicadores. Ejemplos de ello pueden encontrarse en la microplanificación educativa vinculada a unidades territoriales específicas, en donde los indicadores permiten analizar y evaluar aspectos específicos de la política educativa en algunos de sus subsistemas.

Generar información sobre áreas o temáticas relevantes, pero vacantes en términos de su disponibilidad.

A medida que la dinámica de los sistemas educativos evoluciona, se presentan nuevos problemas y con ello se identifican nuevas necesidades de información y de producción de indicadores. Algunos ejemplos de esto son la importancia de producir nuevos indicadores que den cuenta de

las distintas dimensiones de la calidad educativa, en las que los resultados académicos de los estudiantes constituyen solo un aspecto; se ha mencionado también la educación de jóvenes y adultos y las nuevas manifestaciones que ha adoptado; en fin, la información que se produce sobre el financiamiento educativo es un área de vacancia o de desarrollo reciente, pero aún insuficiente para los países de la región.

Evitar el uso de la información producida por los establecimientos educativos para la toma de decisiones relativa a la asignación de recursos a ellos mismos.

Es esta una de las principales razones, aunque no la única, por las cuales la información producida sobre un sistema encuentra inconsistencias y problemas de validez. Las finalidades de un sistema de información, el marco conceptual y el metodológico son etapas de un proceso, a la vez lógico y cronológico, pero profundamente imbricadas entre sí. Si se desvirtúa el sentido de la recolección de la información, es factible que el resto de las etapas del proceso se vean afectadas profundamente, con el consiguiente desperdicio de tiempos, esfuerzos y recursos.

Sobre el análisis de la información

Fortalecer experiencias en la región que promueven la integración de información proveniente de distintos sistemas de indicadores y estudios de evaluación y promover el análisis de la información educativa en forma articulada con aquella proveniente de otras fuentes y otros sectores sociales, para una comprensión integral de los procesos educativos.

Es de destacar que los sistemas de información en la región vienen realizando avances importantes en este aspecto. Mientras el primero de los aspectos remite a un mejor aprovechamiento de los sistemas internos productores de información en un ministerio (oficinas de estadísticas y de evaluación); el segundo reenvía a la importancia de integrar en el análisis la información de contexto que es producida por otros organismos públicos. Cabe resaltar que este punto tiene también implicancias en la organización del proceso de recolección de la información educativa, al definir por ejemplo las unidades de análisis. Si bien es sabido que no siempre resulta sencillo integrar fuentes distintas y estudios complementarios, es importante sostener esfuerzos en ese sentido ya que mejora la interpretación de la información disponible, atendiendo a la complejidad de situaciones de cada sistema educativo.

Fortalecer la generación y difusión de enfoques teóricos en los cuales se sustenta el análisis de la información y profundizar la interpretación de la información con el análisis cualitativo.

Un dato, cualquier dato en sí mismo, no informa sobre algún aspecto de una realidad si no es puesto en relación con enfoques teóricos determinados, que permiten interpretarlo, esto es, dotarlo de sentido. Sin pretensión de extendernos en este punto, porque excede los límites de este trabajo, es necesario aclarar que una mirada atenta a las desigualdades propias de los países de la región, a las formas en que se manifiestan y a las heterogeneidades internas, supone tomar posición en un marco determinado desde el cual esta cuestión se interpreta (que tiene también implicancias en el momento de diseño y definición de un sistema de indicadores). En este sentido, los indicadores proveen insumos al debate sobre el tema.

Por otra parte, la comparación, ya sea en el tiempo para un mismo sistema educativo, como en un contexto, por ejemplo regional, contribuye a atender a la especificidad de la situación educativa de un país o región. En muchas ocasiones, para desarrollar este tipo de interpretaciones es

ineludible complementar la información cuantitativa provista por un conjunto de indicadores con análisis cualitativos provistos por estudios e investigaciones sobre aspectos centrales de un sistema educativo.

Sobre la difusión y el uso de la información

Fortalecer el uso en el planeamiento y la evaluación de las políticas educativas.

Una concepción actualizada del planeamiento educativo, atendiendo tanto al macro-planeamiento como a la micro-planificación, debe basarse en un sistema de información confiable el cual, al aportar evidencia empírica, contribuye a definir los problemas sobre los que las políticas pretenden intervenir, ya sea en el largo, mediano y corto plazo. Es este un insumo fundamental para los responsables políticos y político-técnicos involucrados en los procesos de toma de decisiones, sin que ello implique desconocer el conjunto de otros factores que intervienen en dichos procesos. Asimismo, un sistema de información permite también ejercicios prospectivos en los que se anticipa la evolución del sistema, o se construyen escenarios alternativos como parte del proceso de toma de decisiones.

La articulación entre planificación, gestión y evaluación de un sistema educativo permite incluir distintos elementos para construir conocimiento sobre la evolución de un sistema, así como asegurar la adaptación de las políticas educativas conforme a los desarrollos observados en el sistema y su contexto. Si bien no se trata de un proceso lineal, implica identificar las condiciones que demanden nuevas acciones; inventar, diseñar y desarrollar posibles cursos de acción; elegir entre diferentes alternativas, desarrolladas previamente para solucionar un problema identificado y cuyas consecuencias ya han sido analizadas. En todos estos procesos, la información es una herramienta clave.

Promover diversas acciones de difusión y de capacitación de los distintos usuarios de la información.

Uno de los principales problemas con el uso de los indicadores educativos es, en primer lugar, su falta de uso; en otros casos, los problemas derivados de una interpretación incorrecta de la información que proveen porque, como se afirmó anteriormente, esta supone un conocimiento relativamente complejo de los alcances, las posibilidades y los límites del mismo sistema de indicadores y, al mismo tiempo, un conocimiento de los sistemas educativos sobre los cuales se produce información.

Parte de esta problemática se resuelve pensando en los destinatarios diferenciados de la información y con ello en la elaboración de distintos tipos de informes recurriendo a estilos de comunicación acordes. En ese sentido se constatan esfuerzos interesantes en los países de la región en los últimos años.

Pero, además, el desarrollo de acciones de capacitación en el uso de la información constituye una estrategia válida para contribuir a la sustentabilidad y al mejoramiento de los sistemas de información. Al respecto se debe pensar en distintas "gradaciones" en cuanto a la especialización en el conocimiento de un sistema educativo y en la información producida sobre este, recorriendo un amplio espectro de actores educativos, desde las mismas áreas de los ministerios vinculadas con la gestión de los sistemas educativos, pasando por diferentes niveles de responsabilidad según la organización del gobierno de un sistema educativo. Además, el acercamiento a usuarios "no especializados" de la información con alto impacto social, como los periodistas, por ejemplo, pue-

de resultar interesantes perfiles para organizar acciones específicas destinadas para ellos. Si bien puede contarse en la región con algunas experiencias en esta última línea, son más esporádicas y no se encuentran suficientemente desarrolladas.

Sobre el conjunto del proceso

Profundizar las acciones de formación y capacitación de los recursos humanos en las áreas de estadística y evaluación de Ministerios de Educación de la región.

En la producción de un sistema de información (tomado ahora en la perspectiva más amplia que incluye las etapas ya señaladas), es necesario contar con la intervención de especialistas de diferentes disciplinas y, si bien se han realizado acciones relevantes de formación, es necesario sostenerlas y profundizarlas en el tiempo. La multidisciplinariedad que caracteriza a los equipos responsables de todo el proceso de producción de la información exige además acciones diferenciadas y específicas de formación y, a la vez, implica sostener una visión general e integrada sobre dicho proceso.

Sostener la inversión en el largo plazo para asegurar el mejoramiento y uso de la información.

Generar sistemas de indicadores más complejos, que permitan seguir tendencias en los sistemas educativos, requiere importantes recursos. Como es sabido, las áreas productoras de información son muy demandadas en los Ministerios de Educación; no obstante, no siempre cuentan con los recursos adecuados o quedan sujetas en mayor medida que otras dependencias a la disponibilidad de recursos. Asegurarlos en el largo plazo (en formación de recursos humanos, en equipamiento y su actualización permanente) supone reconocer el importante papel que en esta materia debe jugar el Estado, el cual no puede delegarse en ninguna otra organización social. El apoyo político es de vital importancia para que un sistema de información sea efectivo, tanto en lo referido a la producción como a la utilización de los datos.

A MODO DE CONCLUSIÓN

Resulta claro a partir de los aspectos desarrollados en el presente capítulo que la información sobre un sistema educativo, y un sistema de indicadores generado para tal fin, no puede deslindarse de dos cuestiones centrales.

Por un lado, las preguntas sobre las temáticas relevantes de un sistema educativo, sobre las que se pretende producir información. Aquí el conocimiento de las tendencias, tanto internacionales y regionales como las nacionales, provee una perspectiva importante para la formulación de dichas preguntas. Por ello, la necesidad de actualización de un sistema de indicadores remite a la necesidad de aprehender mejor los nuevos problemas que enfrentan los sistemas educativos, sin que se pierda de perspectiva el análisis de sus tendencias históricas. Pero, por otro lado, resultan clave tanto los enfoques conceptuales como metodológicos, porque remiten a una discusión sobre los supuestos epistemológicos que están detrás de ellos; esto es, qué se conoce de un sistema educativo y cómo se lo conoce; en otros términos, qué porción de la realidad de este representa un sistema de indicadores educativos.

Por último, la producción de información a través de indicadores y su relación con el planeamiento de las políticas educativas constituye una vía para problematizar determinados temas y colocarlos en la agenda pública, que involucra ineludiblemente los más altos niveles políticos de

un Ministerio de Educación. En materia de políticas públicas, se ha afirmado que quien define es quien decide. Esta cuestión permite subrayar entonces la importancia que tienen las oficinas productoras de información, como instancias de apoyo clave para los más altos niveles de responsabilidad política, en la definición de los problemas y en la generación de consenso en torno a ellos, a partir de las pruebas y de los análisis que proveen. Por supuesto, esto supone reconocer también la estrecha, y en muchas ocasiones inseparable, relación entre las dimensiones técnicas y valorativas implicadas en toda decisión de política educativa.

La evaluación de políticas y programas: desafíos para la gestión en educación

Marta Kisilevsky

PRESENTACIÓN

Reconocido como uno de los atributos de *capacidad estatal*, la evaluación –en particular aquella referida a programas y proyectos– ha experimentado un importante crecimiento en las últimas décadas en áreas como salud, asistencia social y educación.

A la complejidad propia del diseño y la gestión de los servicios sociales universales, la política educativa suma algunas particularidades (su institucionalidad, la naturaleza de sus contenidos básicos y los actores que la sostienen) que le confieren atributos propios de los procesos que se suceden en torno al ciclo de la intervención pública.

En un intento por pensar desde esta especificidad, este trabajo pone el foco en la relación entre la evaluación de programas y los procesos decisorios a través de los cuales se diseñan e implementan las políticas y los programas educativos. Se hace un recorrido por sus características básicas, las particularidades derivadas del contexto institucional, sus contribuciones y problemas referidos a los estudios evaluativos, así como los desafíos con que se enfrentan tanto evaluadores como gestores para la utilización de sus resultados.

La reflexión está situada en la interfase que tiene lugar entre los procesos de evaluación y la dinámica particular que asumen el diseño y la implementación de la política educativa; interrogando acerca de las condiciones, aspectos problemáticos y desafíos presentes desde ambas esferas de la relación.

EVALUACIÓN DE PROGRAMAS EN EDUCACIÓN

En su sentido más general, evaluar supone determinar el valor o mérito de algo a partir de establecer criterios y estándares relevantes y realizar posteriormente un análisis adecuado a ellos.

En el ámbito de los programas sociales, tal como señala Mokate (2000, p. 2), la evaluación requiere de la especificación y concertación de expectativas, criterios o estándares y su aplicación a la valoración de las iniciativas. Se trata entonces de procesos que, si bien están situados en el campo del conocimiento científico y técnico, exigen ser *puestos en diálogo* con aquel que se constituye desde la subjetividad de los actores, sus expectativas, necesidades, criterios y valoraciones.

Desde esta perspectiva, algunos autores enfatizan el carácter reflexivo de la evaluación: "tomar distancia de la acción para tomar esta como 'objeto' de análisis, para revisar qué es lo que se está proponiendo o qué se está haciendo o qué se ha hecho, analizar las características de las metodo-

logías de intervención planteadas o desplegadas, y determinar si la orientación es efectivamente hacia la direccionalidad deseable, aclarando cuáles escollos y cuáles facilidades se presentan y qué logros se esperan o se han obtenido" (Nirenberg; 2008, p. 231).

Esta doble pertenencia de los estudios evaluativos de los programas sociales (al campo científico y técnico y a aquel que se constituye desde la subjetividad de los actores) se expresa en los perfiles de sus profesionales, en los modelos de investigación adoptados, en las metodologías de análisis, en las técnicas y en la naturaleza de los datos utilizados y construidos.

Acorde con esta lógica dual, los estudios evaluativos en programas sociales coinciden en la necesidad de combinar los criterios de *confiabilidad y validez* con aquellos que se orientan a abonar a la construcción del *ámbito público* y a la recuperación de la perspectiva de los *actores*[1].

En estas elaboraciones teórico-metodológicas (así como también en sus resultados) existen importantes acuerdos acerca de la importancia que tiene la perspectiva de los actores (sociales y gubernamentales) para la comprensión de los procesos de gestión de los programas. Esta jerarquía conferida a la mirada de los actores se funda en distintas preocupaciones, entre otras: la necesidad de fortalecer con información la democratización de la acción estatal; la importancia de contar con su perspectiva en la construcción de la valoración de los procesos públicos; o bien, por la necesidad de reconstruir la complejidad de la "cascada de la implementación" (en tanto los procesos y los resultados de los programas dependen de los actores que intervienen). Su relevancia se fundamenta desde distintas justificaciones.

Sin embargo, nos interesa destacar que no parece ser suficiente reivindicar la importancia de la mirada de los actores para reconstruir desde un estudio evaluativo la complejidad de la gestión. A nuestro juicio, las dificultades que encuentra parte de las aproximaciones teórico-metodológicas más utilizadas devienen menos de la falta de reconocimiento acerca de la importancia de recuperar la mirada de los actores (que instituye esta "doble pertenencia" a la que hacíamos referencia), que de la propia delimitación de *quiénes son los actores de la gestión de un pro*grama. Y en consecuencia, cuáles son los sujetos de la evaluación.

En este punto, esta pregunta lleva a plantearnos una segunda cuestión que es la relación entre "programa educativo" y "política educativa", sus vínculos recíprocos y sus fronteras. Dicho de otra manera, ¿es posible reducir la complejidad de la "política pública" al artificio del recorte que suponen los "programas" desde sus documentos? Aquí radica –a nuestro juicio– el "nudo" en torno al cual se articulan algunas de las dificultades que encuentran las evaluaciones para alcanzar una relación virtuosa entre "evaluación" y mejora de la "gestión".

Programas y políticas

Aunque existen similitudes entre ambos conceptos y siendo concientes de las limitaciones que encuentra este trabajo para tratar este tema, interesa advertir acerca de ciertos matices que desafían al momento del diseño de la evaluación de programas.

[1] En este registro de la cuestión, podemos citar aportes ya clásicos como los que hicieron décadas atrás E. Cohen y R. Franco (1987), así como otros más recientes, entre los que pueden citarse los trabajos de V. Ruiz, O. Nirenberg y J. Brawerman (2000 y 2003) y de S. Hintze (2001).

Al hablar de "programa" hacemos referencia a acciones que se orientan a objetivos definidos conforme a un determinado diagnóstico de una situación o problema. Ahora bien, a lo largo de su ciclo de vida, los "programas" no se mantienen en "estado puro" ya que, cuando se ponen en marcha, se relacionan con otros programas o con otras formas de organización (servicios, demanda, instancias de apoyo, entre otras).

La relación de los "programas educativos" con la "política educativa" en los distintos niveles (nacional, provincial, departamental o del establecimiento) depende de las características que asumen los procesos de gestión, las condiciones del contexto, las acciones derivadas de otros programas e intervenciones y las capacidades estatales. Este último aspecto resulta clave en tanto las capacidades disponibles colaborarán con el "modelado" del "programa" como "respuesta" a las cuestiones problematizadas en cada nivel. La gestión permite establecer esta conexión entre la "política educativa" y la ingeniería del "programa" ya que, cuando se habla de *gestión de la política educativa*, se está hablando (también) de "programas en acto" (Chiara y Di Virgilio, 2008).

En términos de los interrogantes que orientan nuestro trabajo, estas afirmaciones llevan a recoger en el propio diseño de la evaluación aspectos tales como: la trama de otras intervenciones con las que el programa en cuestión se vincula; las acciones de otros actores (sociales y gubernamentales no directamente vinculados con el programa, pero que pueden incidir –por acción u omisión– en su desarrollo); las estrategias desplegadas en los distintos niveles jurisdiccionales y las capacidades estatales puestas en juego, entre otros.

En suma, esta perspectiva recupera al "programa" en acto en la "política educativa" y supone avanzar en la formulación de interrogantes/organizadores de los estudios evaluativos que estén a la altura de la *complejidad* que desafía a los gestores de los programas, no en tanto instancias aisladas ("discretas" podríamos decir), sino convertidos *en parte de la "red política"*. Desde nuestras preocupaciones, atender a estas necesidades de los gestores resulta relevante en tanto serán estos, en última instancia, quienes podrán retomar (o no) los resultados de los estudios en la toma de decisiones.

Particularidades del campo educativo

Precisadas estas fronteras e identificadas algunas de sus "zonas grises", consideramos necesario indagar en los atributos particulares que asumen los programas en educación y la relevancia que tienen para el desarrollo del sector.

A diferencia de otros campos de la política social, la política educativa presenta algunos rasgos particulares –su diseño institucional, la naturaleza de sus contenidos básicos y sus actores–, que vale la pena destacar a los efectos de ver cómo interpelan a los marcos analíticos de la evaluación.

Con relación al primer aspecto, los atributos de la *matriz institucional* del sector educativo constituyen coordenadas fundamentales para el diseño de la evaluación de programas. Se trata de estructuras organizadas bajo la modalidad de "servicios"[2] con fuertes contenidos jerárquicos en

[2] Una forma tradicional de organizar las funciones sustantivas es creando unidades que se han especializado en la producción o distribución de un servicio que la institución debe entregar a la comunidad (Varas González, 1984).

la estructuración de las relaciones interinstitucionales y en las que intervienen a su vez distintas jurisdicciones con altos niveles de autonomía. En el caso argentino, como en otros países de la región, se trata de una organización federal de gobierno que define competencias diferenciadas y también concurrentes, las que han sido afectadas en las últimas décadas por los procesos de descentralización del sector. Estas particularidades cobran especial relevancia, en tanto se trata también de un sector organizado bajo la lógica de "sistema", cuestión que confiere particular valor a la articulación entre ciclos y niveles.

En segundo lugar, las cuestiones relativas a la enseñanza y el aprendizaje, que están en el foco de la preocupación del sistema educativo, son *problemas complejos*, esto es, difícilmente acotables, porque son multidimensionales (intervienen factores de muy distinta índole: sociales, pedagógicos, institucionales, políticos, culturales, psicológicos) (Bardach, 1997).

Por último, otra de las particularidades tiene que ver con los *actores*, de cuyas intervenciones deviene la *polisemia de sus problemas y de las intervenciones*; en cada caso, los distintos actores le asignan significados, sentido de la urgencia y relevancia diferente. Este atributo está dado precisamente porque coexisten (en su definición, análisis y resolución) en los distintos niveles (nacional, provincial, departamental y del establecimiento) el punto de vista de los alumnos y sus familias y de los demás actores del sistema (maestros, supervisores, directivos, funcionarios y técnicos) que contribuyen también a su definición.

Estas particularidades del sector educación desafían a los estudios evaluativos en los siguientes aspectos, a saber:

- La necesidad de dar cuenta y explicar las articulaciones entre las líneas de acción previstas en los programas con las otras formas de organización con las que se articula matricialmente. Las tensiones que se produzcan entre las intervenciones derivadas de los programas y la organización de los servicios educativos tienen que poder ser explicadas (no solo detectadas).

- La trama institucional que resulta de la descentralización es particularmente heterogénea en los niveles provinciales[3], lo cual supone estructuras organizacionales diferentes, recursos humanos y presupuestarios distintos. Dada esta situación, los efectos de un programa pueden ser unos u otros, según cuáles sean las características que asuman los procesos de gestión, las condiciones del contexto y las capacidades institucionales en cada una de esas matrices para llevarlos adelante. Desde esta perspectiva, el contexto institucional es un factor condicionante que debe ser retomado en el diseño del estudio evaluativo.

- Los estudios evaluativos tienen que poder promover la generación de preguntas que estén a la altura de la complejidad de los problemas. Si bien se reconoce que un estudio evaluativo necesariamente tendrá que operacionalizar las preguntas, deberá hacerlo buscando perder lo menos posible el carácter complejo y multidimensionado del problema.

- Si consideramos que los problemas tienen distintos significados desde los diferentes actores, resulta central que el diseño del estudio evaluativo pueda, recuperando la teoría del cambio que sustenta el programa, ponerla en tensión con las perspectivas que los distintos actores tienen al respecto.

[3] Se corresponde con el nivel estadual de gobierno.

ESTUDIOS EVALUATIVOS: ENTRE LA DECISIÓN Y LA ACCIÓN

Ahora bien, hasta aquí hemos enunciado algunas cuestiones que han buscado conectar la problemática particular de los programas en educación con los atributos necesarios en los estudios evaluativos de programas. Sin embargo, los estudios evaluativos tienen lugar (al igual que los programas) en contextos históricamente determinados y cumplen diversas funciones que hay que considerar en la etapa de su diseño.

La evaluación tiene como función principal la de *tomar decisiones* en torno a una determinada intervención. Sin embargo, las funciones de los estudios evaluativos son diversas; Fernández Ballesteros (1997) señala cuatro posibles, a saber:

1. Conocer acerca de las actuaciones sobre el programa.

2. Justificar las decisiones ya emprendidas.

3. Fundamentar un gasto presupuestario.

4. Contrastar teorías.

Dada la multiplicidad de motivaciones y necesidades que fundan el origen de un estudio evaluativo, parece adquirir particular importancia entonces la tarea del equipo evaluador indagando acerca de "para qué" ha sido encargada la evaluación, esto es, cuáles son sus propósitos tanto explícitos como implícitos. Se trata de un primer avance para "tender puentes" entre la "decisión" y la "acción" desde la evaluación.

Si en el campo de la política pública en general los vínculos entre "decisión" y "acción" están muy mediatizados, en la política educativa este nexo está adicionalmente atravesado por tensiones de distinto orden. Estas tensiones derivan, en parte, de cuestiones ya aludidas relativas a los *atributos de la matriz institucional* del sector y a las *mediaciones* que existen entre la decisión política expresada en la formulación de un programa y la acción efectiva en el nivel de la implementación en los establecimientos y en las aulas.

Las características de las intervenciones así como sus resultados e impactos están fuertemente condicionados por las reglas de juego, los actores y sus capacidades, que se despliegan en cada una de las matrices provinciales. En el caso del sector educación en Argentina post-descentralización, esta afirmación adquiere gran importancia en tanto la diferencia de los contextos institucionales es muy importante y depende de la naturaleza de su forma de organización, la inserción política de la cartera educativa en el gobierno provincial, el perfil de los equipos técnicos, la relevancia que tienen las inversiones derivadas del programa en el contexto del presupuesto provincial, entre otros.

Esta situación pone en cuestión las orientaciones metodológicas más clásicamente utilizadas en evaluación de programas, que suponen la posibilidad de evaluar los "efectos netos" de un programa como si este constituyera una "unidad diferenciada" separada del contexto institucional en el que opera regularmente.

Dada esta situación, resulta necesario avanzar en distintos interrogantes que den cuenta de las tensiones que se suceden en la implementación, tales como: ¿cuáles fueron las estrategias adoptadas para llevar a cabo los cambios propuestos por el programa en cada uno de los niveles?;

¿cuáles son las modalidades de gestión del programa en el nivel provincial?; ¿cómo han incidido las diferentes matrices institucionales y la capacidad de gestión en el desarrollo del programa en cada jurisdicción?; ¿cuáles han sido los modos típicos de constitución de los equipos técnicos provinciales en tanto dispositivo de articulación entre la gestión nacional, la gestión provincial y las escuelas?[4]

La heterogeneidad es entonces parte constitutiva de las situaciones que desafían a los gestores. En consecuencia, dar cuenta de ella es imprescindible para poder generar un conocimiento que sea relevante y significativo, que ilumine el proceso de toma de decisiones de la gestión.

Si toda evaluación compromete una reflexión sistemática y fundada acerca de la acción, esta reflexión no se da en abstracto, sino en el contexto de las relaciones entre los actores de la gestión del programa y de la política educativa. Para que exista una relación virtuosa entre evaluación y acción en la gestión parece necesario que desde la evaluación se indague en las preocupaciones que dan lugar a la evaluación y se formulen –de manera explícita– las preguntas que orientan dicha reflexión.

Sin lugar a dudas, no se trata de una transpolación de las preocupaciones de quien solicita la evaluación a las preguntas que orientarán la indagación. "La formulación de las preguntas constituye un momento crucial del diseño del estudio evaluativo y comprende un *trabajo crítico* del equipo evaluador que debe tener en cuenta las siguientes cuestiones" (DINIECE, 2008)[5]:

a. Las *necesidades y motivaciones* que dieron lugar al requerimiento.

b. El *estado del arte* respecto a la problemática en la que se inscribe el estudio.

c. El *momento* del desarrollo del programa en que la evaluación tendrá lugar.

d. La mirada de *otros actores* vinculados con la implementación del programa y sus preocupaciones.

En consecuencia, la definición de las preguntas de un estudio evaluativo es un proceso de construcción que pone en diálogo las motivaciones de la gestión con los avances en el campo académico. Se trata de un momento inicial pero clave en el proceso de construcción del objeto de la evaluación. Recuperando la preocupación que anima las reflexiones de estas páginas, las preguntas que se formulen deben reunir una serie de atributos, a saber: a) ser *pertinentes* con relación al objeto a evaluar, o sea, el programa; b) ser *relevantes* con relación a las necesidades sociales e institucionales que le dan origen y, por último, c) ser *factibles* de ser respondidas con los recursos y el tiempo disponible.

Desde estos interrogantes se orienta la selección de la estrategia de investigación más adecuada para darle respuesta. Las decisiones en torno al diseño de las evaluaciones responden a criterios tanto políticos (derivados de los requisitos propios de la rendición de cuentas orientada a la construcción de "esfera pública") como técnicos (entre los que se destacan los principios de validez y confiabilidad).

[4] Tomado del estudio evaluativo PIIE (2007).

[5] Se retoman aquí los criterios enunciados en DINIECE, 2008.

Sin embargo, a estas consideraciones generales se suma la necesidad de generar un tipo particular de conocimiento (y sus instancias de validación) que sea capaz de retroalimentar sinérgicamente la gestión en sus niveles críticos.

LOS DESAFÍOS EN TORNO A LA UTILIZACIÓN DE LOS RESULTADOS DE LAS EVALUACIONES DE PROGRAMAS: PROBLEMAS Y ESTRATEGIAS

Desde la experiencia acumulada en estudios evaluativos de programas en educación es posible sistematizar algunas dificultades, problemas y nudos críticos con que se enfrenta la búsqueda de una relación virtuosa entre estudios evaluativos y decisión en la política educativa.

A nuestro juicio, estas dificultades y problemas se sitúan tanto en la esfera de la gestión de la política educativa como en el diseño mismo de los estudios evaluativos.

Con relación a la primera, la *gestión de la política educativa*, cabe destacar los siguientes nudos críticos que, entre otros factores, obstaculizan la relación entre estudios evaluativos y decisión en la política pública:

1. En primer lugar, podemos citar la *falta de incentivos para incorporar la evaluación a la dinámica de la política educativa*. Si bien se han verificado importantes avances respecto a la importancia de la revisión de las acciones públicas y de la rendición de cuentas hacia otras instituciones y hacia la ciudadanía en general, aún las evaluaciones son visualizadas como "auditorías" que, cuando son explícitas respecto a los problemas existentes, pueden ser excesivamente "costosas" para sus gestores. Desde estas representaciones, que forman parte del "sentido común" de los gestores y funcionarios, se constituye y modela una cultura institucional que está poco habituada a recibir imágenes que formulen una crítica. Este registro se ve reforzado, en parte, desde otros actores quienes –desde una noción débil acerca del compromiso de la construcción de la esfera pública– utilizan la información en las pujas intraburocráticas.

2. En segundo lugar, desde la "trastienda" de los estudios evaluativos, pueden consignarse las *debilidades que se encuentran en los diseños de los programas que se van a evaluar*. La forma de organización de la gestión bajo la modalidad de programas requiere del cumplimiento de algunos requisitos que constituyen los componentes del mismo, a saber: la enunciación de un problema y el diagnóstico que lo sostiene, una teoría del cambio explícita que fundamenta la orientación, un conjunto de objetivos que se propone el programa y la población objetivo, un conjunto de actividades y líneas de acción, la explicitación de los recursos con los que cuenta, la cobertura a alcanzar, así como los resultados e impacto. Estos componentes suelen estar presentes de manera explícita en los documentos de los programas, aunque con niveles desiguales de coherencia interna. Sin embargo, el equipo de evaluación puede encontrarse con un conjunto de líneas de trabajo denominadas por sus responsables como "programa", pero que no tenga explicitados la totalidad de estos componentes, o bien los mismos padezcan de ciertas inconsistencias internas (DINIECE, 2008). Frente a esta dificultad, el equipo evaluador deberá reconstruir el objeto que se pretende evaluar: se trata de una tarea inicial necesaria para hacer posible el diseño del estudio evaluativo.

3. Un tercer factor que desafía la relación entre evaluación y decisión política es la *complejidad institucional que mediatiza la teoría del cambio que sustenta el programa*. Aunque, tal como ya enunciáramos, la matriz institucional (diversa, conforme a las distintas provincias, depar-

tamentos y establecimientos en que el programa despliega sus acciones) es parte insoslayable y constitutiva de la gestión de la política educativa, es indudable que se trata de un factor que relativiza los resultados de los estudios evaluativos. Si bien los programas están organizados conforme a una determinada teoría del cambio, las características de las matrices institucionales (diversas según las jurisdicciones) mediatizan la relación causal entre las acciones previstas en un programa y los efectos esperados.

En la otra esfera, existe un conjunto de factores sobre los cuales los responsables de las evaluaciones tienen mayor margen de gobernabilidad, y se trata de aquellos derivados del *diseño mismo de los estudios evaluativos*:

1. Uno de los factores relativos al diseño alude a la *composición del equipo de evaluación y la dinámica de trabajo asumida con el equipo a cargo del programa*. La conformación del equipo evaluador es un momento fundacional en el proceso evaluativo. Si bien algunos perfiles básicos deben "timonear" la formulación del estudio, los perfiles deben definirse conforme a los requerimientos del diseño del proyecto y deben comprender perfiles más sustantivos (relativos a la problemática educativa de que se trate) y otros más orientados a la gestión de políticas.

   La capacidad comunicativa, de diálogo y problematización conjunta con los actores es una competencia de fundamental importancia para garantizar un tratamiento virtuoso de la relación entre los resultados y las decisiones de los gestores. Esta competencia deberá ponerse en juego en *reuniones y encuentros de trabajo* con decisores/solicitantes de la evaluación y en *entrevistas en detalle y conversaciones* con actores vinculados con el programa; estos espacios son tanto o más importantes que los mismos informes que resultan del estudio.

2. El segundo factor se refiere a cómo el diseño adoptado para el estudio puede *captar las diferencias conforme a estas mediaciones institucionales y los atributos de capacidad de los actores*. En contextos federales como el argentino resulta fundamental la posibilidad de recuperar los estudios institucionales y de política educativa en el nivel provincial para generar información que resulte adecuada a la diversidad de contextos de implementación en que los programas se despliegan.

3. Un tercer factor tiene que ver con la *escala de los resultados*. Si bien las posibilidades presupuestarias inciden considerablemente en estos casos, la representatividad de los datos debería poder estar a escala de la intervención (nacional, si el programa es nacional, o provincial, si se corresponde con una iniciativa de esta jurisdicción); esta decisión es central para construir interlocución y legitimidad frente a los interlocutores de la gestión.

4. Por último, un factor que incide y es complementario de los anteriores tiene que ver con la *naturaleza del diseño del estudio*. Los estudios solo cuantitativos o cualitativos pueden perder o bien potencial analítico o bien capacidad de reconstrucción de escala. Para dar cuenta de la complejidad de los fenómenos bajo estudio se requiere una adecuada triangulación de fuentes e integración de datos cuantitativos con la perspectiva de los actores y los registros documentales.

Se trata de una tarea compleja, que exige rigurosidad en las competencias específicamente educativas, pertinencia metodológica en los aspectos institucionales del diseño y capacidades de comunicación para el intercambio. Desde nuestra experiencia parece ser –aunque difícil– la mejor apuesta aportar virtuosamente a informar desde la evaluación la acción pública.

Currículo y evaluación estandarizada: colaboración o tensión

Elena Martín

INTRODUCCIÓN

En el primer capítulo de este mismo libro, Alejandro Tiana apunta algunos de los principales debates que se mantienen en el campo de la evaluación. Uno de ellos es precisamente el de la relación entre las políticas curriculares y las políticas de evaluación. Dicho en otros términos, la relación entre el proceso de definición de las intenciones educativas y el proceso de comprobar en qué medida estas intenciones se van alcanzando.

El objetivo de este capítulo es profundizar en la naturaleza de esta relación, presentando argumentos que ayuden a desentrañar si se trata de procesos que se potencian entre sí; si, por el contrario, se corre el riesgo de que las evaluaciones estandarizadas del rendimiento académico de los estudiantes interfieran en las funciones del currículo; o si, como sucede en muchas ocasiones, depende de cómo se pongan en cada caso en marcha ambos procesos.

Para realizar este análisis, en el primer apartado se recuerda el papel que el currículo desempeña en la calidad de la enseñanza y se apuntan las cuatro características que mayor relación tienen, a nuestro modo de ver, con las evaluaciones estandarizadas. A las funciones de estas evaluaciones se dedica precisamente el segundo epígrafe. En el tercero se presentan los principales riesgos que pueden derivarse de determinadas prácticas evaluadoras y, finalmente, el capítulo acaba proponiendo algunas medidas que ayudarían a que evaluación y currículo constituyeran dos pilares de la calidad educativa.

LA DEFINICIÓN DE LAS INTENCIONES EDUCATIVAS

La función del currículo es definir las intenciones educativas que un sistema escolar tiene para con su alumnado. Es decir, seleccionar, de entre todas las capacidades que un ser humano necesita desarrollar para desenvolverse en su entorno como un ciudadano con plenos derechos y deberes, aquellas que no están aseguradas por otros contextos educativos y requieren, por tanto, de la intervención educativa intencional que supone la escuela. El currículo es, pues, un proyecto social en la medida en que se propone contribuir a un tipo determinado de persona. Esto significa entre otras cosas que su definición no puede ser ajena a opciones ideológicas y que variará en determinados aspectos dependiendo del grupo social de referencia.

Acertar en la selección de lo imprescindible resulta, por otra parte, una tarea sumamente difícil, ya que implica discernir aquellos aprendizajes propios de la educación básica obligatoria, sin los cuales alguien se enfrenta al resto de su vida en inferioridad de condiciones. No se trata, pues, de todo lo que desearíamos que aprendieran nuestros jóvenes, sino de lo que les pondría en riesgo

de no conseguir apropiarse de ellos. Asumir esta diferencia entre *lo básico imprescindible* y *lo básico deseable* (Coll y Martín, 2006a) representa ya una primera dificultad, a la que se añaden otras muchas, como el rápido cambio del conocimiento, el peso de la tradición o la diversidad de concepciones sobre el valor de los distintos saberes.

La tercera característica del currículo que nos interesa destacar, además de su naturaleza social y por ello ideológica y su función de selección de lo esencial, es la dificultad que implica su desarrollo al perseguir que todos los alumnos aprendan lo básico común y, a ser posible mucho más, y para ello prestar ayudas específicas y diferenciadas que se ajusten a la diversidad de formas de ser y de aprender de todos los estudiantes. Armonizar lo común y lo diverso es difícil, pero también imprescindible si se quiere que el currículo definido llegue realmente a convertirse en aprendizajes relevantes. El debate sobre las ventajas y desventajas de un currículo abierto, así como las que se producen en torno a las decisiones de descentralización curricular, cobran sentido desde esta perspectiva.

Finalmente, el aspecto que quizá más relevante resulte desde el punto de vista de las relaciones entre currículo y evaluación estandarizada sea el hecho de que lo que se pretende ayudar a adquirir a través de la intervención escolar son competencias, es decir, aprendizajes complejos, que además remiten a ámbitos muy variados del desarrollo humano.

El énfasis actual en las competencias no se trata de un mero cambio terminológico, aunque de no apoyarlo convenientemente pudiera correrse el riesgo de acabar siéndolo. Es cierto también que no es una propuesta totalmente novedosa, ya que presenta claras afinidades con los modelos centrados en capacidades. No obstante, supone un énfasis en determinados puntos nucleares del aprendizaje que, de llevarse a la práctica, podrían contribuir de forma muy notable a la mejora de la enseñanza. En concreto, conviene recordar los cuatro rasgos esenciales de la adquisición de una competencia.

El primero, su funcionalidad. Ser competente implica ser capaz de utilizar un conocimiento para solucionar un problema. Los saberes inertes, que no sirven para actuar, no constituyen aprendizajes significativos. La segunda característica remite a la implicación del conjunto de la persona en esta movilización. Ser competente supone activar conjuntamente recursos cognitivos y emocionales: hay que entender, pero hay que querer, hay que mantener el esfuerzo, hay que superar la ansiedad de la incertidumbre y de los errores que forman parte de cualquier aprendizaje. La integración de varios elementos se manifiesta también, como tercera característica, en el uso conjunto de tipos de contenido que es preciso coordinar para actuar de forma competente. Cualquier problema relevante requiere conocimientos de tipo conceptual, pero también procedimientos y desde luego formas de aproximarse al problema que remiten a determinados valores y actitudes. El último atributo de la competencia es su adecuación al contexto, el uso funcional de los conocimientos solo producirá una actuación competente en la medida en que responda a las demandas específicas del contexto. Se es más competente cuanta mayor sea la variedad de contextos en los que se actúa con éxito, pero no se aprenden competencias generales descontextualizadas, sino que se desarrolla la capacidad de analizar lo específico del contexto y ajustar la respuesta a ello. La experiencia en muchos contextos va permitiendo ampliar el rango de actuación y generalizar lo aprendido.

La caracterización de lo que significa una competencia ilustra con claridad la complejidad de su naturaleza, que a su vez aumenta cuando se destaca el amplio rango en el que debe desarrollarse.

Desde la acertada propuesta de los pilares del informe Délors, no ha cesado de hacerse hincapié en que las competencias remiten a campos que trascienden lo que ha venido siendo la tradición más academicista. Sirva como ejemplo el énfasis que Cox (2008) hace en la dificultad de alcanzar las capacidades de creatividad, innovación y flexibilidad que el informe de la "New Commission on the Skills of the American Workforce" de Estados Unidos, publicado en 2006, propone como metas esenciales, si se continúa con las políticas actuales.

En síntesis, el currículo en el que se plasman las intenciones educativas responde a una serie de funciones: adecuarse a las demandas sociales propias del grupo social para el que se define y a sus opciones ideológicas mayoritarias; seleccionar los aprendizajes básicos imprescindibles; permitir la atención a la diversidad para garantizar que estos aprendizajes básicos los adquieren todos los alumnos, aunque ello exija distintos procesos de enseñanza; contribuir al aprendizaje de competencias complejas y que cubran los diversos ámbitos del desarrollo. Los procesos de evaluación estandarizada deberían corresponderse con estas características. A valorar en qué medida esto sucede dedicaremos el tercer apartado, pero antes conviene recordar las funciones que por su parte se atribuyen a la evaluación del rendimiento académico.

EL PAPEL DE LAS PRUEBAS DE EVALUACIÓN ESTANDARIZADA

La función fundamental de este tipo de pruebas es comprobar si el sistema escolar está cumpliendo el encargo recogido en las intenciones educativas que se expresan en el currículo. Es responsabilidad de las administraciones educativas, como instrumentos del Estado, garantizar el derecho a una educación de calidad de todos los estudiantes. Para supervisar el cumplimiento de esta obligación es preciso contar con instrumentos de evaluación de los logros educativos, entre los que se encuentran estas pruebas.

Esta función de comprobación de los niveles de aprendizaje puede utilizarse con una finalidad diagnóstica y formativa o puede ponerse al servicio para fines de mayor impacto asociadas a procesos de acreditación de los estudiantes o de las propias instituciones educativas.

Su carácter común tiene, entre otros, el objetivo de que en esta comprobación se preste especial atención a la equidad de los resultados. Una vez superado el concepto de equidad como mero acceso a un puesto escolar e incluso como la igualdad de oferta educativa, el indicador más adecuado es la igualdad de resultados. Es decir, que estos no estén asociados a determinadas características del alumno, como género, etnia o nivel sociocultural. Las pruebas de evaluación estandarizada, cuando tienen en cuenta estas variables en sus análisis, ilustran el nivel de equidad de los sistemas educativos.

Algunos autores atribuyen también a las pruebas la capacidad de ayudar a que los docentes acuerden entre sí el contenido esencial de sus evaluaciones y, en algunos casos, también la forma de llevarlas a cabo. La enorme disparidad en las prácticas evaluativas de los docentes es un problema comprobado empíricamente (Grisay, 2002). Cuando los resultados de las evaluaciones estandarizadas se hacen llegar a los centros adecuadamente, como explicaremos con más detalle en el último apartado, pueden servir para suscitar interesantes discusiones sobre la forma de evaluar en el aula.

Esas discusiones pueden llegar a que la revisión sea más amplia y abarque el conjunto del currículo. Los datos de rendimiento de los alumnos sin duda ofrecen una información interesante sobre

los aprendizajes realizados por estos, en comparación con los del resto de las instituciones, que puede desencadenar procesos de innovación curricular.

A estas funciones algunos autores añaden la de definir el currículo. Por ejemplo, en el interesante documento que PREAL ha publicado recientemente (Ravela, Arregui, Valverde, Wolfe, Ferrer, Martínez Rico, Aylwin y Wolff, 2008) se recoge esta función:

> "El desarrollo de un sistema nacional de evaluación estandarizada puede obligar a realizar una discusión informada sobre qué aspectos del currículo formal son exigibles a todos los estudiantes y a *definir con claridad* qué es lo que todos los estudiantes deberían haber aprendido al final de cada ciclo educativo" (p. 8. El énfasis en cursiva es nuestro).

En este mismo sentido se considera en muchos casos que las evaluaciones estandarizadas del rendimiento académico no solo definen lo básico del currículo, sino que deben guiar el conjunto de las políticas curriculares (Solomon, 2002). Ferrer y Arregui (2003) lo expresan así en otro de los documentos editados por PREAL:

> "Este es también el momento de *revisar el contexto de política curricular*: en qué momento de implantación de los programas curriculares se encuentran el país y sus regiones; cuáles son las dificultades que evidencian las pruebas nacionales, los informes de supervisión regionales, las investigaciones académicas y la prueba internacional misma para decidir qué tipo de mejoras (si no cuáles exactamente) se pueden recomendar a las autoridades nacionales y subnacionales de educación" (p. 24).

Estas dos últimas afirmaciones son, a nuestro juicio, las más polémicas y por ello queremos analizarlas en el siguiente apartado, junto con otros posibles riesgos de la evaluación estandarizada.

LOS RIESGOS PARA EL CURRÍCULO DE LAS EVALUACIONES ESTANDARIZADAS DE LOGRO

Es habitual encontrar en los textos que tratan este tema la idea de que el problema de la evaluación estandarizada es, en todo caso, que no es un instrumento suficiente para garantizar todas las funciones de control, rendición de cuentas y mejora que se atribuyen a los procesos evaluativos. Nosotros coincidimos sin duda en estas limitaciones y señalaremos las que se refieren a la valoración del cumplimiento de las funciones del currículo, pero nos atreveríamos a afirmar que en algunos casos la evaluación estandarizada implica riesgos que no se reducen a omisiones, sino a posibles errores tanto en el plano de las políticas de la administración como del funcionamiento de los centros.

Por lo que se refiere a las ausencias, lo más llamativo es la dificultad y, en algunos casos, imposibilidad de evaluar determinadas competencias. Incluso los sistemas de evaluación más evolucionados, como pueden ser las pruebas de PISA, el TIMMS o las que realiza el LLECE, que se han expuesto en la primera parte del libro, se limitan a evaluar algunos conocimientos de matemáticas, de ciencias de la naturaleza y de lengua –habitualmente comprensión lectora–. Esto supone una gran limitación desde la perspectiva de los ámbitos del saber. Los aprendizajes vinculados con lo artístico, lo social, lo emocional o lo moral están prácticamente ausentes. Este reduccionismo a menudo se refleja también en una primacía de los conocimientos conceptuales y, aunque en menor medida, de los procedimientos frente a las actitudes y los valores.

Esta selección, que con acierto Martínez Rizo (2005) denomina "muestreo de contenidos", a veces se explica por razones técnicas, pero el argumento de mayor peso remite ante todo a prioridades

académicas que con demasiada frecuencia no son coherentes con lo que por su parte proclaman los documentos internacionales de mayor calado sobre las competencias básicas (OCDE, 2005; Comisión Europea, 2004). La creatividad, la innovación y la flexibilidad, que se citaban más arriba como capacidades esenciales para el mundo actual y futuro, no tienen una mayor vinculación con los conocimientos habitualmente evaluados que con el resto de los saberes.

Nadie sensato negaría el carácter básico del uso funcional de la lengua, tanto escrita como oral, la capacidad de ordenar y representarse el mundo a partir de sus características matemáticas, ni la necesidad urgente de la alfabetización científica. El problema es que también cualquier persona lúcida compartiría que es imprescindible ser capaz de enjuiciar la moralidad del comportamiento propio y ajeno, de conmoverse con aquello que lo merece, de controlar la ansiedad que genera la incertidumbre o de sentirse responsable del bienestar físico y emocional de los que nos rodean e incluso de aquellos a los que no conocemos personalmente. Sin embargo, estas competencias apenas aparecen en las pruebas estandarizadas de evaluación.

Incluso si pudiéramos llegar a justificarlo por la complejidad técnica que supone evaluar estos aprendizajes –argumento que podría cuestionarse en algunos casos–, no por ello hay que dejar de indicar que las evaluaciones mandan desde este punto de vista un mensaje muy pernicioso acerca de las intenciones educativas. Establecen un "metanivel" de lo básico al seleccionar, de lo que ya ha definido el currículo como esencial, aquello que parece preocupar de verdad y que por ello se evalúa para el conjunto del sistema.

No es fácil que este mensaje que tan nítidamente llega a los docentes quiera luego hacerse compatible con que el profesorado otorgue igual importancia al resto de los conocimientos definidos como igualmente básicos. Puede lograrse, pero no es una forma de actuar que contribuya al cambio radical que la selección del qué enseñar necesita en este momento. Por el contrario, introduce una tensión que pocas veces se resuelve con el adecuado equilibrio.

La evaluación de competencias, por otra parte, exigiría en sentido estricto que las pruebas incluyeran algunos tipos de tareas que no es posible, al menos por el momento, llevar a cabo en un modelo estandarizado. Si bien se está avanzando mucho con la incorporación de preguntas abiertas, de respuesta construida, todavía estamos lejos de lo que un seguimiento continuado del comportamiento de un alumno por parte de un docente puede informar acerca de su nivel de competencia. Es cierto que las pruebas internacionales citadas, así como otras nacionales, muestran una forma de evaluar competencias en sí misma modélica, muy por encima del nivel medio de los colegios e institutos, pero esto no significa que estemos siendo capaces de realizar evaluaciones *auténticas* de todos los aprendizajes que remiten a las competencias básicas, aumentando con ello la invisibilidad de los conocimientos que no quedan recogidos en esas pruebas.

El carácter estandarizado de las evaluaciones plantea un segundo riesgo para el currículo: la dificultad de ser coherentes con el nivel de diversidad al que deben dar respuesta las intenciones educativas. La descentralización administrativa del currículo y la autonomía de las instituciones escolares –dos procesos relacionados, pero no idénticos, como señala Tiana en este mismo volumen– se han mostrado potentes factores de calidad y, sin embargo, muchas veces entran en contradicción con la prioridad de lo común que caracteriza a las evaluaciones estandarizadas. Por otra parte, se trata no solo de la homogeneidad de contenido de los aprendizajes que se incluyen en las pruebas, sino también de las formas de demostrar los aprendizajes. Como muchas veces sucede, está claro que quienes resuelven las pruebas han aprendido; sin embargo, no lo está

tanto que muchos de los que tienen ciertas dificultades con ellas tengan niveles de competencia tan bajos como los que se derivan de los resultados. Las pruebas reflejan un determinado modelo de cómo enfrentarse al conocimiento, el más habitual. Este modelo es sin duda valioso, pero no agota todas las formas de usar el conocimiento para resolver problemas. Esta disonancia se hace más notoria cuando se analizan los resultados de grupos de estudiantes pertenecientes a culturas minoritarias.

La diversidad es no solo de estudiantes, sino también de escuelas. Los análisis que se realizan tomando el centro como unidad de calidad son especialmente interesantes, pero hay que tomarlos con mucha prudencia. La autonomía curricular implica que las escuelas establecen prioridades en las intenciones educativas sin que ello tenga que hacer peligrar lo básico. Pero además entre los distintos momentos de aplicación de las evaluaciones se producen muchos cambios que hacen difícil interpretar adecuadamente los niveles de rendimiento y, sobre todo, las causas a las que podrían atribuirse. Como señalan Linn, Baker y Betebenner (2002), es preciso tener en cuenta la "volatilidad" de las escuelas: de un año a otro cambian las cohortes de alumnos, la plantilla de profesores, el equipo directivo…, lo que hace muy difícil atribuir los resultados únicamente a las decisiones curriculares que el centro haya podido tomar. Cuando a esto se une una mala política de difusión y apoyo a las escuelas en la interpretación de los resultados, la imagen que estas pueden hacerse no supone en la mayor parte de los casos un buen punto de partida para los planes de mejora.

Finalmente, querríamos abordar un problema de mayor calado, ya que se refiere a una crítica de carácter epistemológico. Como hemos expuesto en detalle en otro lugar (Coll y Martín, 2006b), quizá la mayor debilidad del modelo que propone una relación directa entre evaluación estandarizada y currículo sea el reduccionismo epistemológico que, a nuestro juicio, comporta el supuesto de que es posible remontarse sin mayor dificultad de las informaciones sobre los resultados de rendimiento del alumnado a los factores que los explican, de que es posible realizar inferencias sobre las causas del rendimiento directamente a partir de la medida del mismo. Y lo que es aún más complejo y discutible, el supuesto de que las causas a las que se atribuyen los resultados de rendimiento se sitúan inequívocamente en el ámbito del currículo. Actualmente sabemos que el proceso mediante el cual las intenciones educativas establecidas en un currículo o en una relación de estándares acaban concretándose en unas determinadas experiencias docentes y de aprendizaje para el profesorado y el alumnado en los centros educativos y en las aulas, y a través de ellas en unos determinados niveles de rendimiento, es sumamente complejo y que en él intervienen e inciden, orientándolo en uno u otro sentido, multitud de factores. Intentar recorrer y reconstruir, siquiera parcialmente, el proceso en sentido inverso con el fin de identificar y valorar los factores curriculares directamente implicados en el mismo es una tarea ciertamente mucho más compleja y cargada de incertidumbres de lo que sugiere una lectura ingenua del esquema que estamos comentando. No es una tarea imposible, pero sí costosa y de resultados previsiblemente discutibles. El reduccionismo epistemológico que impregna esa racionalidad se manifiesta de nuevo con claridad si atendemos a su vertiente más propositiva y proactiva, es decir, a la utilización de las informaciones sobre el rendimiento del alumnado *con el fin de tomar decisiones* en el ámbito del currículo. Incluso suponiendo que las precauciones adoptadas nos permitan recorrer y reconstruir el proceso en sentido inverso, llegando así a formular conjeturas razonables sobre la incidencia de los factores y procesos relativos al currículo en los niveles de rendimiento observados, ¿cómo podemos derivar lógicamente a partir de ellas propuestas concretas de cambio curricular? Entre los resultados de las evaluaciones de rendi-

miento, su interpretación y su proyección en propuestas concretas de qué se debe modificar en el currículo, hay un salto epistemológico que solo puede justificarse recurriendo a elementos ajenos en sentido estricto al esquema que estamos analizando. El elemento clave de los procesos de toma de decisiones implicados en la conducción de los cambios curriculares no son los resultados de las evaluaciones de rendimiento, sino el filtro interpretativo utilizado para derivar a partir de ellos propuestas concretas de acción. Y el ingrediente principal de este filtro son precisamente las intenciones educativas, de cuya consecución el rendimiento evaluado es en definitiva un indicador, y que constituyen, como antes señalábamos, el referente último para valorar la idoneidad de un determinado currículo. En síntesis, las evaluaciones de rendimiento pueden proporcionar y proporcionan informaciones sumamente útiles sobre el grado de consecución de las intenciones educativas, pero no serían la fuente de la que surgen estas intenciones ni tampoco el instrumento adecuado para su legitimación.

Sería útil hacer una revisión de las reformas curriculares que se han realizado en los países de América Latina, en la que se analizara en qué medida las decisiones tomadas se han fundamentado en resultados de evaluaciones de rendimiento. Cuando en nuestro caso llevamos a cabo este estudio en las reformas curriculares españolas ocurridas entre 1990 y 2005, no encontramos ninguna conexión. Los cambios curriculares se basaban en opciones teóricas de carácter piscopedagógico y, ante todo, en posiciones ideológicas que impregnaban tanto el currículo como el conjunto de las reformas del sistema educativo. En un interesante trabajo que Ferrer (2004) llevó a cabo acerca de las reformas curriculares de Argentina, Chile, Colombia y Perú, también señala que la repercusión de los datos sobre rendimiento incide poco en la toma de decisiones de política educativa. Sin embargo, su impacto en la innovación en las escuelas parece ser mayor:

> "Es también en Chile donde los datos sobre resultados de aprendizaje parecen tener mayor impacto en la toma de decisiones informadas para el diseño de políticas. Prueba de esto es el uso de los resultados para la incorporación de escuelas seleccionadas a ciertos programas focalizados, o para la asignación de incentivos institucionales y docentes. En Argentina, los datos no han sido utilizados mayormente para la toma de decisiones, aunque sí para el diseño de estudios de seguimiento a la implementación curricular. En Colombia, el uso de los resultados tampoco ha informado decisiones puntuales de política, aunque las autoridades del sector se refieren a ellos frecuentemente para fundamentar sus opiniones y juicios sobre la calidad y necesidades del sistema.
>
> El uso de los resultados para fines pedagógicos parece ser más extendido que el uso como herramienta de responsabilización política o escolar. Todos los países realizan algún tipo de devolución de los resultados que involucra el análisis de algunos ítems tomados de las pruebas, y luego presentan una serie de recomendaciones pedagógicas para solucionar los problemas más críticos. En los casos de Chile y Bogotá, donde las pruebas son censales, los resultados se devuelven a todas las escuelas, y docentes y directores pueden ver cuáles son algunos de los contenidos en que sus propios alumnos tuvieron mayor dificultad, y así tomar medidas remediales" (p. 172).

Todos los datos apuntan a que las evaluaciones estandarizadas de rendimiento por sí solas no garantizan una mejora del diseño ni del desarrollo curricular. Sin embargo, cuando se ponen en marcha determinadas medidas, los resultados sobre el rendimiento de los estudiantes pueden ser un potente instrumento de cambio. En el último apartado de este capítulo analizaremos cuáles son las actuaciones de las que, a nuestro juicio, dependen que evaluación y currículo se potencien mutuamente. Destacaremos únicamente las que consideramos más relevantes de entre las muchas que se han venido señalando en la bibliografía más reciente (ver, por ejemplo, Ravela y otros, 2008).

LA COLABORACIÓN ENTRE CURRÍCULO Y EVALUACIÓN

Si la propuesta que se ha defendido en el apartado anterior se considera acertada, sería lógico admitir también que, para que los resultados de la evaluación puedan provocar mejoras curriculares, es preciso llevar a cabo procesos de interpretación cuyo acierto dependerá en gran medida del nivel de conocimiento que quienes realizan la interpretación tengan de los múltiples factores que pueden estar de hecho mediando entre las intenciones educativas definidas y el desarrollo curricular real. Esta condición se cumple con mucha más facilidad cuando el análisis de los resultados se lleva a cabo en el nivel del centro que cuando se hace para el conjunto del sistema educativo. La escuela se convierte así en la unidad clave para esta conexión. Por otra parte, esto puede ayudar a resolver la tensión entre el carácter necesariamente común de las pruebas estandarizadas y las diferencias que la autonomía curricular produce de hecho en los aprendizajes de las distintas escuelas.

Resulta fundamental, por lo tanto, hacer llegar a cada escuela los datos de la forma más adecuada. Es decir, lo antes posible, presentando los resultados con claridad y ofreciendo las comparaciones pertinentes, apoyando con expertos y ofreciendo la colaboración que sea necesaria por parte de la Administración para apoyar los cambios que se propongan.

En el interesante trabajo que Benveniste (2002) realizó analizando el caso de Uruguay se destaca que el positivo impacto que las evaluaciones de rendimiento tuvieron sobre la educación de este país se debió, entre otras cosas, al sentimiento de confianza logrado por la unidad de evaluación responsable de las evaluaciones entre los cuadros de maestros que, en opinión de este autor, podía ser atribuido a cuatro factores:

1. Estricta confidencialidad de los resultados de las pruebas.

2. Rápida devolución de los resultados de las pruebas a las autoridades de las escuelas.

3. Contextualización de los puntajes de las pruebas según la situación sociocultural.

4. Abstención de responsabilizar directamente a los maestros por los logros académicos.

En el trabajo con los centros debería incluirse también a las familias. La opinión de los padres sobre las pruebas de evaluación es fundamental, y no siempre es fácil transmitir a la opinión pública los resultados de estos procesos. La interpretación que los docentes pueden ayudar a hacer a las familias puede resultar muy útil para ir creando una cultura de la evaluación, además obviamente de para ayudarles a conocer con rigor la situación en la que se encuentra su hijo y el resto del alumnado de la escuela.

Por otra parte, es conveniente recordar que quienes diseñan y desarrollan en último término el currículo son los docentes. Ellos son, pues, la pieza clave en la interpretación de los resultados de la evaluación y la elaboración y desarrollo de los planes de mejora curricular. La formación del profesorado, centrada en este caso en el análisis de los resultados de las pruebas de evaluación, resulta una pieza clave. En el caso de Uruguay se organizaron seminarios largos de dos y tres meses, que ayudaron al trabajo interno de cada escuela.

El trabajo de los docentes está mediado, además de por la formación, por los materiales curriculares, que se convierten así en un factor determinante en la gestión del cambio. Pocos docentes acceden al currículo leyendo directamente las normas y también son escasos los que elaboran sus

propios materiales didácticos. Los libros de texto y los materiales curriculares en general, en los distintos soportes en los que en este momento se presentan, desempeñan un papel fundamental, sobre todo si se utilizan en el marco de acciones de formación inicial y en servicio.

Schmidt y Prawatt (2006) realizaron un estudio en 37 países en el que comprobaron que los libros de texto eran el factor que más correlacionaba con la coherencia y consistencia del currículo, ejerciendo una influencia mucho mayor que el control administrativo centralizado.

El trabajo más directo con las escuelas y los profesores es el medio más potente para generar la discusión sobre el currículo que, en su caso, pudiera llevar a cambios en las intenciones educativas. No obstante, hay que ser realistas y ajustar las expectativas. Como Goh, Cousins y Elliot (2006) señalan en su artículo, apenas hay cultura de utilizar los datos de evaluación entre los equipos de las instituciones docentes. Por ello, las administraciones deberían planificar de forma meticulosa estas actuaciones y mantener el apoyo de forma constante.

Por otra parte, es importante coordinar estos procesos de reflexión con otros foros de debate más amplios en que puedan participar expertos así como las propias administraciones. En la revisión que Ferrer y Arregui (2003) realizaron sobre el impacto de las evaluaciones internacionales en América Latina se señala que debe aprovecharse el momento en que se entregan los resultados de las pruebas para generar discurso innovador y para convocar foros de interpretación en los distintos niveles del sistema educativo.

No obstante, a nuestro juicio, las actuaciones que se han señalado hasta aquí deben completarse con otras más estructurales y duraderas. Probablemente, una de las mejores formas para promover la colaboración entre currículo y evaluación estandarizada del rendimiento académico es contar con instituciones nacionales que se dediquen a ambos temas y que trabajen en absoluta coordinación. La existencia de centros de desarrollo curricular y de institutos de evaluación es imprescindible para la calidad de la enseñanza, pero la clave reside en su trabajo conjunto. Las pruebas deben ser realizadas por expertos en currículo y estos deben analizar posteriormente los resultados para analizar las consecuencias de los resultados. El reciente documento de PREAL sobre las evaluaciones que América Latina necesita (Ravela y otros, 2008) señala precisamente esta medida como una de las necesarias.

Cuando se dan la mayoría de estas condiciones, se comprueba que los resultados de las evaluaciones de rendimiento ayudan a la mejora del currículo y de la calidad de la educación en general. La opinión de uno de los docentes entrevistados en el trabajo de Benveniste (2000) sobre el caso de Uruguay refleja la idea que, a nuestro modo de ver, es la más importante: que la difusión de los resultados de la evaluación sirva para *abrir caminos*:

> "Esta evaluación masiva en nuestro aprendizaje ha permitido ofrecer […], a nivel del sistema educativo, algunas herramientas que permitieran corregir en parte esta situación de bajo rendimiento de los niños mediante algún sistema de actualización de los docentes y proponiendo algunas estrategias útiles, sobre todo a nivel de psicología, de lengua y matemáticas. Yo creo que, desde este punto de vista, tuvo una función formativa muy importante a nivel del país. Permitió que muchos docentes se conectaran [con sus alumnos] porque muchos sabían que las cosas estaban mal, pero no estaba muy clara la explicación del porqué. *Sirvió para buscar caminos*" (Entrevista UGN10, p. 21. El énfasis en cursiva es nuestro).

La evaluación de centros: riesgos y posibilidades

Leda Badilla

INTRODUCCIÓN

Como se ha señalado en la introducción del libro, uno de los problemas en el campo de la evaluación educativa es la poca atención que se ha puesto a la evaluación de los centros escolares contrastada con otros ámbitos, como son los logros del aprendizaje. Es en el centro educativo donde convergen funciones evaluativas relacionadas con logros académicos, pero también con las familias, los docentes, los estudiantes, la evaluación misma y el sistema educativo en general.

Evaluar el centro educativo de una manera más integrada responde a una mejor forma de comprender la organización escolar pero, sobre todo, a mejorar los procesos que ocurren dentro y fuera de él. El objetivo del capítulo es analizar los antecedentes y los supuestos teóricos en los que se fundamenta este tipo de evaluación, presentar algunos modelos de evaluación de centros que ilustran estos principios y destacar las dificultades que podrían explicar la escasa presencia de la evaluación de centros y algunas sugerencias para superarlas.

ANTECEDENTES

Desde los años cincuenta, la corriente educativa asociada a la igualdad de oportunidades trajo consigo una serie de informes que alertaron sobre el fracaso escolar y las desigualdades educativas de origen social y económico[1].

En los años sesenta prevaleció la lógica del desarrollo expansionista que agregó poco análisis a la relación existente entre políticas educativas, acciones y resultados. Posteriormente se hizo evidente que, a pesar de la extensión de la cobertura, el fracaso escolar persistía. La atención se dirigió entonces a comprender las finalidades del centro escolar y cómo este es afectado por diferentes factores. El Informe Coleman (Coleman *et al.*, 1966) concluyó que el contexto familiar está fuertemente relacionado con el resultado académico de los estudiantes, lo que se explicaba en ese momento por la herencia cultural, el nivel educativo de los padres y sus expectativas educativas y, por otro, por la escasa relación entre los recursos disponibles en el centro y el rendimiento de los estudiantes. Diez años más tarde, Coleman volvió a analizar sus datos, considerando otros estudios, y afirmó que los efectos de la escuela en los estudiantes eran más importantes que lo señala-

[1] En los años cincuenta, el informe "Abandono temprano" y el "Informe Crowter". En los años sesenta, los informes "Gilly" en Francia, "Coleman" en Estados Unidos de Norteamérica, así como los estudios de Roeder y colaboradores en Alemania.

do por él mismo diez años antes (Casassus, 2003). Las reacciones ante los resultados expresados en el informe no se hicieron esperar: se generó más investigación para profundizar y determinar el papel de la escuela en el rendimiento de los estudiantes, así como nuevas propuestas de investigación e intervención.

En 1983 se publicó el informe "Una nación en riesgo", que criticó el desempeño de la educación norteamericana, lo que provocó un giro en la política educativa de ese país orientándola hacia la "calidad", noción que reemplazó los temas "igualdad" y "democracia", que inspiraron muchos de los estudios de las dos décadas anteriores.

La calidad de la educación se puso sobre el tapete en las políticas educativas de la mayoría de los países y se vinculó a diferentes componentes de la educación, como el currículo, los docentes, la gestión escolar y la supervisión, entre otros.

La evaluación de los centros educativos también se abordó desde la lógica organizativa; por ello, algunos modelos desarrollados en las últimas décadas se ubican como paradigmas organizativos. Tal es el caso del modelo de las escuelas eficaces, el cultural-pluralista o los modelos para la calidad, entre otros (Martín-Moreno, 2007).

El movimiento "Mejora de la eficacia escolar"

Este movimiento surge como una de las nuevas propuestas que introduce variables de proceso atribuidas a la institución escolar e identifica factores de eficacia y calidad de la educación. Se caracteriza porque reúne dos tendencias: una vinculada a la eficacia escolar, y otra a la mejora de la escuela. Aunque sus enfoques teóricos y metodológicos son diferentes, ambas colaboran para que los centros mejoren (Muñoz-Repiso *et al*., 2000). Esta aproximación es una forma de comprender los datos evaluativos articulándolos a elementos teóricos. De este modo se abandona la idea de que los aspectos socioculturales o intelectuales serían los que expliquen linealmente el fracaso de los estudiantes. Los estudios de escuelas eficaces se ocupan de cuestiones sustantivas, como la investigación empírica, los procesos de cambio educativo y organizativo, la eficacia de programas específicos, las prácticas evaluativas, la participación de la dirección, la autonomía y el trabajo docente, entre otros (Báez, 1994).

Los estudios de eficacia escolar (Muñoz-Repiso *et al*., 2000):

- Consideran una visión sistémica del centro.
- Incluyen cuatro niveles de análisis: estudiantes (aptitudes, habilidades, perseverancia, calidad de la docencia), aula (calidad de la docencia, incentivos y tiempo de dedicación), escuela (metas significativas compartidas, funcionamiento académico diario, coordinaciones entre participantes, organización para el aprendizaje) y contexto (comunidad local, administraciones regional y central).
- Trabajan con hipótesis y resultados empíricos.
- Utilizan la investigación previa de forma flexible para definir nuevos temas de investigación, y orientan la intervención.
- Emplean el modelo CIPP (contexto, entrada, proceso y producto) con algunas variaciones.

El movimiento sobre mejora de la escuela despegó con la utilización de nuevos materiales didácticos (década de los sesenta), posteriormente incursionó en la intervención en las prácticas educativas y, por último, se identifica como la gestión del cambio.

Los estudios de eficacia escolar, por su parte, dan especial importancia al clima del centro educativo como agente decisivo, ya que determina el comportamiento de otros factores como las tasas de éxito, la repetición y la satisfacción global. Entre los factores de eficacia se mencionan el liderazgo profesional, los objetivos compartidos, el clima positivo de trabajo, el refuerzo positivo a los estudiantes, la calidad del proceso enseñanza-aprendizaje que se lleva a cabo en el aula.

Afortunadamente, ambos enfoques –escuelas eficaces y mejora de la escuela– han convergido en un modelo conjunto –"Modelo de mejora de la eficacia escolar"– (Muñoz-Repiso *et al.*, 2000), en el que se identifican siete elementos que se relacionan con el éxito de los programas de mejora. Ellos son la cultura escolar, el papel de la dirección, la organización escolar, la planificación, los procesos de enseñanza-aprendizaje, el estímulo externo y las experiencias previas.

Otras experiencias

Se esperaría que la evaluación de centros fuera una práctica común, con algún grado de sistematización, pero por lo general no es así. Si bien existen proyectos educativos en los que se incluyen centros con determinadas características, no constituyen evaluaciones propiamente dichas. Las tendencias a poner en práctica ciertos proyectos o iniciativas surgen cada cierto tiempo. En la mayoría de los países de la región, los temas de estos proyectos se designan con expresiones como "Escuelas de calidad", "Escuelas prioritarias", "Escuelas líderes", "Colegios Bandera Azul", entre otras. Me referiré a tres experiencias que están documentadas.

Una de ellas la presenta el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (UNESCO, 2002); se trata de un estudio cualitativo realizado en establecimientos de siete países de la región[2], escogidos por sus resultados destacables, específicamente en matemática, en el Primer Estudio Regional Comparativo. El calificativo de destacables se aplica a esas escuelas porque el rendimiento de sus estudiantes estuvo por encima de lo que se esperaría dado el nivel educativo de sus padres. El desempeño positivo de estas escuelas, a pesar de las condiciones de adversidad en que se insertan, se atribuye principalmente a dos factores: la gestión escolar y la práctica pedagógica en el aula. En el primer caso, se considera el ambiente de aprendizaje, el papel de alumnos y padres, la organización escolar para la gestión, el establecimiento de relaciones emocionales positivas, normas para la convivencia social, la estabilidad laboral y capacidad de liderazgo tanto de directores como de docentes y la eficacia en la organización, entre otros. En lo referente a la práctica pedagógica del aula, estas escuelas han mostrado innovación en las estrategias y principios para promover el aprendizaje; el espacio es un recurso que se utiliza de diversas formas, no hay una perspectiva pedagógica única; el educador se muestra como orientador; los métodos didácticos son muy variados; se utilizan tecnologías de la información; existe preocupación por el desarrollo integral de los estudiantes entre otros.

Otro proyecto tuvo lugar en Argentina, donde se desarrolló una experiencia evaluativa en algunas jurisdicciones de la ciudad de Buenos Aires (Rodríguez y Puiggrós, 2002). Su propósito fue mejo-

[2] Los países participantes fueron: Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba y Venezuela.

rar los aprendizajes por medio de vínculos democráticos al interior de las prácticas pedagógicas. En este caso, el equipo docente asumió una responsabilidad central, en el sentido de definir los aspectos que tomaría en cuenta la evaluación, así como el análisis de la información y las líneas de acción que se debían seguir una vez conocidos los resultados. Los aspectos se refieren a procesos de aprendizaje sistemático y no sistemático, el clima institucional, los recursos materiales y la atención a diversas problemáticas sociales. Los resultados se vincularon, a su vez, a un proyecto permanente de capacitación, así como a acuerdos interinstitucionales y de participación no tradicional de las familias en la escuela. También se puso en práctica un mapa curricular consensuado; se hizo visible lo que se enseña y lo que se evalúa, así como la necesidad de contar con espacios institucionales para definir acuerdos de trabajo y la puesta en práctica de una variedad de climas de trabajo y organización, entre otros.

La tercera experiencia la constituye el programa Escuelas de Calidad puesto en práctica en México. La evaluación externa realizada a los centros de educación preescolar que participaron en dicho programa entre 2003 y 2006 muestra que esos centros elaboraron estrategias de mejora y perfeccionamiento en sus planes de transformación escolar. En esas escuelas se atiende a las familias de escasos recursos económicos; se fortalece la gestión; se pone en práctica el liderazgo de las directoras de los centros; se crean condiciones para el trabajo colegiado; se abren las instituciones a la comunidad, además de que los resultados de la evaluación externa se hacen públicos (Myers *et al*., 2006).

SUPUESTOS TEÓRICOS DE LA EVALUACIÓN DE CENTROS

El enfoque de evaluación de centros que, a nuestro juicio, resulta más interesante es el que entiende que la práctica evaluativa no puede centrarse solo en el estudiante. Llegar al juicio evaluativo requiere, entre otras tareas, reconocer la posición de los sujetos que participan, las individualidades y contingencias en relación con alguna situación que les involucra en el propio centro y en su contexto cercano. Este juicio reflexivo (Žižek, 1998) induce a pensamientos más profundos sobre los hechos de la vida cotidiana y la naturaleza de las cosas. Por su parte, los centros escolares se caracterizan por la evaluación de lo particular (Simons, 1999). Lo particular se establece como la búsqueda del sentido del razonamiento práctico, que no es algo que se aprende primero de la teoría y luego en la práctica individual, sino que se confronta con una necesidad, lo que nos permite construir ese juicio informado (Carr, 1996) deseable cuando nos encontramos frente a una evaluación seria.

Partimos de la realidad de que las transformaciones escolares no se efectúan aisladamente, sino que se relacionan con otras, situadas en contextos diversos que de una u otra forma convergen en la vida del centro. El centro escolar lo concebimos como la unidad donde es posible el cambio. Por ello la evaluación no la podemos asumir si no es articulada a los docentes, a la cotidianidad del aula, a los estudiantes, a los padres y madres y a los equipos docentes. El propósito que fundamenta este enfoque de evaluación es formativo, es decir, de mejoramiento.

Fundamentada en la larga trayectoria de evaluar centros escolares, Simons (1999) concluye que es la escuela, y no el aula, la unidad principal de cambio, es a ella a la que se dirigen los esfuerzos de desarrollo. Del mismo modo para Estefanía y López (2003) "es a nivel de centro educativo donde tiene sentido hablar de implicación personal [...], de unidad básica de cambio o del reto de pequeñas innovaciones como referencias para un cambio más global" (p. 17).

El centro se evalúa con el propósito de producir información que sea pertinente y particular, por un lado, para formular propuestas que mejoren la capacidad formativa de este y, por otro, para alimentar las políticas educativas, ya que

> " se va configurando como uno de los pilares básicos de la evaluación del propio sistema, y seguramente como el aspecto fundamental del mismo, por la influencia que tiene sobre el resto de los aspectos que configuran e interactúan conjuntamente" (Estefanía y López, 2003, p. 43).

Cada centro va construyendo una identidad propia que lo distingue de otros centros, según las circunstancias en las que se desempeñe y las áreas curriculares o extracurriculares que lo caractericen.

La metodología y las técnicas para evaluar los centros suelen ser más amplias, adaptables y flexibles, comparadas con los modelos clásicos de corte cuantitativista, debido al contexto cultural de procedencia de los diferentes participantes.

La utilización de diferentes técnicas de investigación (cualitativas o cuantitativa o ambas) ha significado, para sus respectivos defensores, algunas críticas. Se continúa fortaleciendo los aspectos técnicos, no así los teóricos o metodológicos[3], lo que podría constituirse en un riesgo, ya que con frecuencia las técnicas se vinculan más a las conclusiones de los estudios, en lugar de integrarse con los conceptos, escenarios y creencias. Se recomienda mejorar las técnicas de recolección de datos, así como asegurar más validez (House, 1999).

¿Buscamos centros competentes?

En la actualidad, las tendencias en la evaluación educativa se dirigen a las competencias, entendidas estas como la actuación eficaz en situaciones determinadas, que se apoyan en los conocimientos adquiridos y en otros recursos cognitivos (Condemarín y Medina, 2000).

La competencia bien entendida no es solo que las personas se apropien de la información, la procesen y la logren aplicar, sino que integren valores sociales, habilidades interpersonales y actitudes. Además se requiere que dicha integración se movilice para enfrentar situaciones complejas que presentan analogías en su estructura (Perrenoud, 1999) y el contexto en el cual la competencia se materializa (de Macedo, 2002). Si lo que prima es la eficiencia técnica, la estandarización o el funcionamiento, estaríamos frente a procesos que no son las competencias "bien entendidas". Aceptamos una evaluación por competencias que se base en juicios reflexivos porque, como ya se ha señalado, estos permiten la articulación de la práctica educativa con sus determinantes y con los estados contingentes. La evaluación por competencias contextualiza, lo que no significa que evaluar competencias sea un asunto sencillo. Son los diferentes actores de la comunidad educativa quienes con sus prácticas, problemas o expectativas sugieren si efectivamente las competencias de los profesionales que conducen los procesos de la educación cuentan con las competencias que permitan valorar la diversidad de desempeños esperados.

[3] Desde mi práctica evaluativa, considero la noción de pluralismo cognitivo y eclecticismo (Alonso, 1998). El primero valora críticamente la diversidad teórica, ya que no todas las teorías ni todos los métodos son utilizables en general, sino que cada teoría y cada metodología se adecúa a los conocimientos de las realidades sociales, desde donde los investigadores nos posicionamos. Mientras que desde el eclecticismo se añaden elementos teóricos dispersos e inconexos para tratar de explicar una realidad social.

¿Qué hace el centro educativo como unidad frente a las condiciones de adversidad social a las que se ven sometidos niños y jóvenes? ¿Cómo abordan sus problemas socioeducativos si es que efectivamente se abordan? ¿Qué tipo de competencias requiere un centro en un contexto dado? Es esa puesta en práctica de una serie de competencias la que nos permite evaluar las producciones y aportaciones de la proyección profesional docente.

La evaluación del centro, entonces, conduce a identificar problemas que se dimensionan, en términos de las percepciones y motivaciones de los diferentes actores en este proceso, desde un nosotros y no desde un aquellos, desde las diferentes temporalidades[4].

El proceso evaluativo es complejo desde el momento en que se aborde dentro de una perspectiva humano-social. Diferentes actores educativos interactúan con el centro desde sus identidades, concepciones de mundo, valores, problemas y condición socioeconómica. En consecuencia, la unidad "centro educativo" se construye permanentemente.

Un centro competente es aquel que:

- Conoce las particularidades que pueden estar afectando el aprendizaje de sus estudiantes.
- Logra difundir y poner en práctica un proyecto educativo consensuado y evaluable.
- Toma conciencia, acepta y actúa de acuerdo con la variabilidad de las relaciones personales.
- Adecúa el currículo oficial a las condiciones específicas del centro: contexto socioeconómico y cultural, características particulares de estudiantes, docentes, padres, madres, padres, municipalidades, entre otros.
- Logra que sus miembros participen en actividades específicas cuyos resultados siempre sean de mejora y alcanzables.
- Logra que la autoridad se practique de formas múltiples y democráticas.
- Conoce y respeta los procesos de comunicación y coordinación.
- Maneja la incertidumbre y las contingencias.
- Tiene capacidad para identificar conflictos y buscar soluciones.
- Logra apertura hacia el entorno y viceversa.
- Hace atractivo y adecuado el espacio físico.
- Respeta la diversidad sociocultural de estudiantes, sus familias y los docentes.

Llegar a construir un centro competente requiere de acciones evaluativas internas, externas o ambas, y un acompañamiento profesional.

[4] Ardoino (2000) destaca la presencia de las metáforas de la *observación* (inspección) y del *espacio* (panóptica) en las prácticas evaluadoras controladoras, más que la *escucha* y, por consiguiente, el *tempo*, este último en términos de procesos no acabados.

Evaluación interna y evaluación externa

Tanto la evaluación interna como la externa cuentan con métodos y procedimientos propios, que responden a sus propósitos. Lo importante de destacar es que ambas modalidades se articulan entre sí cuando las instituciones optan por realizar dos tipos de evaluación. La mayoría de los países europeos han incorporado de manera sistemática ambas modalidades en los centros de enseñanza obligatoria (Eurydice, 2004).

La evaluación externa puede conducirla un solo evaluador, como generalista, o bien varios evaluadores. Puede responder a objetivos de supervisión de normativas o bien al desarrollo de la calidad escolar. En la evaluación interna, el centro se evalúa él mismo desde sus docentes, dirección, estudiantes, familias, juntas escolares. Esta participación asegura los procesos de mejora, porque los mismos actores identifican los problemas, lo que los compromete también en las soluciones.

Los contenidos de las evaluaciones se vinculan de uno u otro modo a los alumnos, los docentes, las responsabilidades que asumen los centros, la gestión y la administración, el funcionamiento global de la institución y los resultados de los estudiantes.

Las dos perspectivas evaluativas pueden desarrollarse de forma paralela; a veces los temas por evaluar son los mismos.

La *evaluación interna* es un proceso reflexivo que permite a la institución autorreconocerse en sus fortalezas y debilidades y proceder a subsanar estas últimas casi de forma inmediata. Es necesario reconocer que cada evaluación interna es única. Por su parte, la *evaluación externa* la llevan a cabo especialistas, es más imparcial. Ambas modalidades cuentan con procedimientos e instrumentos propios, no excluyentes. Otra modalidad derivada de las dos anteriores es la "semiexterna", que se lleva a cabo desde el mismo centro, pero con la participación de especialistas externos (Estefanía y López, 2003). Lo que sí está claro es que "una escuela necesita cierto grado de autonomía para ejercer su responsabilidad profesional y perfeccionarse desde dentro" (Simons, 1999, p. 240).

Elementos que constituyen la evaluación integral de centros

La tendencia generalizada de contar con una serie de indicadores suele perder de vista las especificidades y particularidades a las que hiciera referencia en las secciones precedentes. En todo caso, los indicadores deben ser consistentes con el enfoque metodológico que orienta la evaluación, que a su vez es coherente con la teoría pedagógico-educativa que sirve de sustrato al objeto de la evaluación.

El análisis global inicial de una evaluación y su consecuente apertura hacia la realidad nos compromete con las preguntas ¿por qué se va a evaluar este centro?, ¿para qué evaluarlo?, ¿a quién van dirigidos los resultados de la evaluación?, ¿cuáles son las posibles consecuencias? Los elementos más comunes sobre los que trabaja la evaluación de centros son los siguientes:

Estudiantes

- Características y condiciones socioeconómicas.
- Contexto socioafectivo.

– Habilidades para el aprendizaje.

– Habilidades en diferentes áreas curriculares.

Docentes

– Clima de trabajo.

– Relación con estudiantes y familias.

– Atención a la diversidad.

Procesos

– Del centro, que se articulan a diferentes actores.

– Del aula.

Familia

– Valoración de la gestión del centro.

– Vinculación a docentes.

– Contexto sociocultural.

– Participación e interacción con el centro.

Las fuentes de información más relevantes para la evaluación externa o interna son:

– Informes anteriores de los que disponga el centro.

– Estudiantes.

– Docentes.

– Padres y madres.

– Director del centro.

– Planes de estudios.

Las fuentes por sí solas no brindarían más que información. Lo que interesa es pasar a mejores elaboraciones analíticas y, consecuentemente, a mejores valoraciones reflexivas. Un plan de evaluación plantea requisitos como son la formación de equipos, la elaboración de materiales, la sensibilización sobre la cultura evaluadora tanto a nivel administrativo como en los propios centros. Cabe aquí replantear las funciones que tradicionalmente ha desempeñado la supervisión escolar y considerarla dentro de este plan de mejora (Estefanía y López, 2003).

¿Quiénes juzgan el quehacer del centro?

Diferentes actores involucrados en la práctica educativa, conjuntamente con los evaluadores, tienen derecho a participar en la construcción de juicios reflexivos sobre el desempeño del centro.

Es claro que los docentes constituyen un grupo fuerte para juzgar lo que el centro hace; también es cierto que los juicios serán más consistentes y equitativos en la medida en la que participen diferentes interesados. La evaluación interna la llevarán a cabo los mismos centros educativos con un comité coordinador del proceso. Es importante que participen diferentes representantes de la comunidad educativa, que estén vinculados al centro o tengan algún tipo de cercanía. Para ello contarán con apoyo de la administración de la educación en sus órganos asesores o normativos.

La evaluación externa la conducirán las respectivas supervisiones educativas, conjuntamente con las direcciones especializadas del nivel central o regional y diferentes representantes de la comunidad educativa, como pueden ser padres y madres de familia u otros órganos especializados.

Otra modalidad interesante de evaluación de centros la constituye el programa "padres y docentes" (PTA), que funciona en algunos estados norteamericanos. En él se involucra a la escuela con los padres y la familia bajo seis criterios: la estrecha comunicación entre la escuela y los padres, y estos y la escuela; el desarrollo de habilidades de apoyo en la labor de ser padres; la participación directa en el aprendizaje de sus hijos; el voluntariado de los padres en la escuela; la plena participación de los padres (voz y voto) en las decisiones que afectan a los niños y las familias, así como la utilización de los recursos comunitarios para fortalecer escuelas, familias y aprendizajes. Cada criterio posee sus respectivos instrumentos para la evaluación, y los resultados cuentan formalmente para la toma de decisiones (National PTA, 2000).

Una vez analizados los resultados, tanto en la evaluación externa como la interna, se pasará a la elaboración de planes de mejora, con plazos consensuados entre las partes participantes. Es decir, la información requiere comprensión e interpretación, ya sea a lo interno del centro como a lo externo, para el mejoramiento que se busca.

Es recomendable que la coordinación de los equipos externos o internos de evaluación la ejerza una persona con formación en el campo de la evaluación. Esta persona es la que organiza las visitas, la administración de instrumentos, entrevistas, etc., pero sobre todo asesora en aspectos teórico-metodológicos a lo largo del proceso evaluativo. El plan de mejoramiento se concreta en proyectos educativos de centro, a los que haré referencia más adelante.

Instrumentos utilizados en la evaluación

En el proceso evaluativo se debe prestar atención a la elaboración de instrumentos. Aunque esta parezca una tarea simple, es quizá la más compleja si se siguen los procedimientos técnicos que demanda la validez de los resultados. No basta con "pasar" a expertos los instrumentos ya elaborados por un grupo o una persona para que los juzguen; es indispensable clarificar los constructos que se valorarán y las consiguientes ponderaciones que diferentes especialistas realicen para cada uno de ellos. Los pilotajes y sus resultados permiten cuidar aún más rigurosamente las facetas de la validez.

Por lo general, se utilizan escalas tipo Likert, listas de cotejo, guía de tópicos para los grupos focales u otros instrumentos específicos para determinadas técnicas evaluativas. Si el enfoque evaluativo aborda perspectivas más cualitativas, también es indispensable la utilización de procedimientos de triangulación, previamente definidos, que garanticen la credibilidad de los resultados.

El informe de evaluación

Es una de las fases más delicadas cuando se desarrollan evaluaciones serias. Todo el trabajo de campo, la sistematización de la información producida puede verse afectada si no se lleva a cabo una "adecuada" devolución de resultados. La característica más importante es la coherencia entre todos los elementos contenidos en el informe.

La información producida y analizada se devuelve al centro. Se discute en grupo para tomar decisiones. Se requiere un informe que esté disponible y pueda ser leído previamente por los interesados. Las personas involucradas socializan la información para:

- Comprender los diferentes problemas y abordarlos considerando sus causas y posibles consecuencias.
- Concretar acciones de mejora. Cada mejora se clarifica y sus respectivas acciones se planifican, esto no acaba hasta tanto el cambio se consolide en el centro escolar.
- Expresar los problemas en un lenguaje comprensible para todos.

Los cambios para la mejora del centro escolar se concretan, como lo señaláramos, en un proyecto institucional, que es conocido como proyecto educativo de centro (PEC).

El proyecto educativo de centro

El proyecto educativo de centro (PEC) se ha constituido en una herramienta que permite la planificación, la gestión y el consiguiente desarrollo del centro. Los componentes del PEC se pueden resumir en: el análisis institucional y su horizonte; los objetivos; las normas para la convivencia y funcionamiento; los aspectos curriculares internos y externos, y las evaluaciones de proceso y producto que demanden los elementos citados (Castillero *et al*., 2002).

El PEC de cada escuela es particular y diferente al de otros centros en lo que concierne a valores, rasgos, dinámicas externas e internas, procesos, prácticas e intenciones educativas. Diferencias que se traducen en lo que el centro "es" y hacia dónde se dirige.

Las normas para el funcionamiento del centro contemplan la participación de la comunidad educativa, la gestión democrática, la información y la comunicación, así como las relaciones con el entorno.

Los aspectos curriculares son los que hacen más evidente al PEC en una realidad concreta. El centro se vuelve competente para atender a las particularidades de sus estudiantes en lo que se refiere a estilos de aprendizaje; condiciones socioemocionales y socioeconómicas, entre otras. La ejecución y evaluación del PEC, por lo general, comprende un análisis situacional del centro; un subproyecto curricular; un plan de desarrollo y un plan operativo. El subproyecto curricular considera elementos de formación cultural, intelectual, científica, tecnológica, socioafectiva, lingüística, ambiental (Castillero *et al*., 2002).

Cabe señalar que, en paralelo al PEC, trabaja la supervisión escolar, que a lo largo de los años ha ido cambiando hacia estrategias más democratizantes, cooperativas, orientadoras y flexibles.

ALGUNOS MODELOS PARA LA EVALUACIÓN DE CENTROS

En la evaluación de centros con frecuencia se observa la utilización del modelo CIPP (contexto-entrada-proceso-producto) con algunas variantes en su diseño o aplicación. Lo fundamental del

modelo se mantiene en sus tres propósitos: servir de guía para tomar decisiones, proporcionar información para los responsables de la institución o programas y comprender los fenómenos en estudio (Stufflebeam y Shinkfield, 1993).

En la evaluación del contexto se valora el estado global del objeto, se examinan metas y prioridades.

La evaluación de entrada posibilita un programa que efectúe los cambios necesarios, en este caso, en un centro escolar. Es preciso contar con un grupo asesor que guíe los diferentes procedimientos a poner en práctica dependiendo de las aplicaciones que se busquen según objetivos propuestos.

En la evaluación de proceso se comprueba la realización de un plan, proporciona información a interesados.

La evaluación de producto tiene como propósito interpretar y juzgar los logros de un programa. Las diferentes fases evaluativas citadas aplican métodos y técnicas para producir y analizar la información requerida.

En general, los modelos evaluativos siguen las siguientes fases: consultas a diferentes grupos para producir la información necesaria; análisis de la información y definición de estrategias para el centro con la participación de los órganos o personas cuyo juicio sea necesario. La devolución de la información requiere un lenguaje comprensible para los diferentes participantes. Esto con el fin de que efectivamente la evaluación conduzca a las mejoras necesarias en el centro.

Basándose en el modelo CIPP se ha desarrollado un interesante modelo de evaluación de centros por parte del Instituto de Evaluación y Asesoramiento Educativo (IDEA)[5] con un enfoque que contempla: el contexto (sociocultural del alumno y del centro); la familia; el nivel inicial de los estudiantes en diferentes asignaturas; los procesos del centro y del aula, y los resultados en cada uno de ellos. El modelo establece líneas diferentes de relaciones: cada nivel influye en el más próximo, algunos niveles influyen en otros más complejos y los diferentes niveles interactúan entre ellos (Marchesi y Martín, 2002).

Toda la investigación evaluativa ha puesto de manifiesto que el nivel socioculutural de la familia del estudiante es una variable estrechamente relacionada con el rendimiento escolar, como también lo es, aunque en menor medida, el contexto sociocultural del centro escolar El alumno no puede despojarse de sus condicionantes sociales ni del centro en el que se forma; por ello se hace necesario "integrar estos aspectos diversos incorporándolos en un marco interpretativo coherente" (Tiana, 2001, p. 61).

En el modelo de IDEA se evalúan: los procesos de centro relativos a la valoración del equipo directivo; la actitud ante la participación y la diversidad del alumnado; la relación con los alumnos; la relación entre los profesores; el trabajo por departamento; el conocimiento del funcionamiento del centro; la satisfacción de las condiciones de trabajo.

En los procesos de aula se incluyen la atención a la diversidad; la amplitud del currículo, la evaluación, el clima del aula. Se evalúan las habilidades metacognitivas, las actitudes y valores, y las diferentes áreas o asignaturas curriculares.

[5] IDEA cuenta con su sede en España y con proyectos específicos en algunos países latinoamericanos.

Por otra parte, se evalúa el rendimiento del alumnado en las áreas curriculares, pero también en actitudes, habilidades metacognitivas y desarrollo emocional. La evaluación responde a un enfoque longitudinal que permite analizar la evolución del alumnado a lo largo de su escolaridad y relacionarla con el resto de los factores señalados.

POSIBILIDADES Y RIESGOS DE LA EVALUACIÓN DE CENTROS EDUCATIVOS

En este último apartado expondremos brevemente, a modo de síntesis, las potencialidades que la evaluación de centros ofrece como elemento de mejora de la calidad de la educación, así como los riesgos que puede entrañar si no se realiza con las suficientes garantías.

Las posibilidades que brinda la evaluación

La función educativa de la evaluación facilita la reflexión en los diferentes niveles del sistema, lo que a su vez permite conocer, reconocer y comprender el centro educativo en sus capacidades para dirigir otras prácticas educativas posibles. No podríamos cambiar lo que no comprendemos.

Si las evaluaciones responden a necesidades de los centros, es probable que se crea en ellas porque reflejan la realidad que estos viven y se aumenten así las posibilidades de que la calidad mejore; lo mismo sucedería con las comprensiones e interpretaciones de las prácticas educativo-pedagógicas que reporta la evaluación.

Si la evaluación utiliza informaciones producidas por diferentes personas (estudiantes, padres y madres, docentes, directores), estará dando credibilidad a los juicios que surjan, más aún si esas informaciones están contextualizadas.

Un diseño flexible de evaluación se basa en diferentes fuentes de observación y aborda tanto procesos como productos, se preocupa de lo general como de lo específico.

Un centro que se evalúa, interna o externamente, cuenta con una mayor probabilidad de que mejoren las prácticas educativas, la gestión del centro en beneficio de la calidad, la planificación de acciones, los aprendizajes de docentes, padres y madres, estudiantes, directores. Se llegan a reconocer valores que antes no fue posible identificar.

La evaluación se desarrolla en función de cada centro y sus circunstancias. Las características comunes de las escuelas nos hacen observarlas como homogéneas; sin embargo, cada una presenta rasgos propios con implicaciones curriculares y extracurriculares variadas.

La adecuada devolución de la información producida por la evaluación, así como las motivaciones que muevan al centro, asegurarían la mejora, tanto para el mismo centro como para las familias y los estudiantes.

Los riesgos a los que nos podríamos enfrentar

Los riesgos se vinculan con la propia naturaleza de la evaluación como un campo del conocimiento controversial. Durante años y aún hoy se le ha querido dar un carácter técnico-administrativo e instrumental, desplazando su sustento teórico y axiológico, así como sus implicaciones políticas y sociales. La evaluación cataliza los cambios, pero ella por sí sola no constituye el cambio.

El valorar un centro cuando no hay investigación básica suficiente, o cuando los recursos no permiten conducir la evaluación, o bien cuando el centro se resiste, puede hacer más difícil, compleja o larga la evaluación.

Las expectativas, tras una evaluación, son altas. Para que los cambios tengan sostenibilidad es necesaria la comunicación abierta, la cooperación y la participación comprometida, sin perder de vista otra serie de factores que permanentemente influyen en el centro, como la movilidad del personal, las bajas motivaciones, rutinas, temor o resistencia ante los cambios que podrían tener lugar.

A veces los establecimientos u órganos educativos no cuentan con profesionales con competencias para conducir los planes de mejora.

Una vez evaluado el centro, es necesario asegurar la permanencia del equipo docente en la institución, ya que es en él en quien recae principalmente la construcción y puesta en práctica de un plan de mejoramiento o proyecto educativo de centro. La responsabilidad no es solo de la dirección del centro.

El manejo prudente de la comunicación de los resultados redundará en la credibilidad de la evaluación.

Si los aspectos de la organización escolar se compaginan con los procesos académicos derivados de la evaluación, las propuestas para la mejora serán coherentes y coordinadas y los elementos contingentes tanto como las incertidumbres se manejarán con mayor seguridad.

Podríamos concluir que el centro escolar media entre el aula y el sistema y, como consecuencia, es el núcleo y objeto de la evaluación; sin embargo, la evaluación del sistema suele limitarse a la consideración de los resultados del logro académico de los alumnos. Pero las escuelas necesitan conocerse desde dentro, además de ser valoradas desde fuera, para determinar sus prioridades y las acciones afirmativas que les permitan la mejora global.

La idea de mejorar la escuela se vincula de forma directa a políticas educativas que la sitúen en primer plano para favorecer su calidad.

Lo anterior, a la vez, nos remite a fomentar una cultura evaluativa en los centros, que se traduzca en prácticas que generen aprendizajes permanentes para toda la red que constituye la comunidad educativa, y para procesos de cultura organizativa, como el clima laboral, el liderazgo, la planificación, que se consolidan en el proyecto educativo de centro.

La escuela no puede producir cambios si estos no se llegan a insertar en un marco de políticas bien definidas y si no se hacen acompañar de profesionales con las competencias necesarias para llevar adelante los planes de mejora y el apoyo continuo.

La evaluación del desempeño docente para el desarrollo de las competencias profesionales

Pedro Ravela

LA EVALUACIÓN COMO HERRAMIENTA DE CAMBIO EN LA EDUCACIÓN

La evaluación tiene un enorme potencial como herramienta para promover y sostener el cambio en la educación, en la medida en que se la utilice bien. En la última década y media, la mayor parte de los esfuerzos, tanto a nivel internacional como nacional, han estado puestos en el desarrollo de sistemas de evaluación estandarizada de los aprendizajes de los estudiantes. Esto ha significado un aporte para la educación, en la medida en que ha permitido hacer visible en qué grado los estudiantes están aprendiendo lo que se espera que aprendan.

Sin embargo, como herramienta aislada y sin articulación con otros instrumentos de política educativa, la evaluación difícilmente pueda producir cambios (GTEE, 2008).

El problema se agrava cuando se intenta organizar el conjunto de la política educativa sobre la base exclusiva de la evaluación estandarizada, por ejemplo, estableciendo premios económicos en función de los resultados de la misma. Se ignora de esta manera que las evaluaciones tienen diseños específicos, apropiados para ciertos fines, pero no para otros. Se desconoce además la necesidad ineludible de establecer un conjunto articulado de políticas educativas.

Y, sobre todo, se olvida que en el sistema educativo ocurren cotidianamente múltiples formas de evaluación, en general realizadas de manera inapropiada, que si fuesen realizadas de forma adecuada tendrían un potencial de cambio enorme. Algunas de ellas son evidentes, como la evaluación de los aprendizajes de los estudiantes que realizan los propios docentes. Otras pasan fácilmente desapercibidas, a pesar de su importancia, como es el caso de las evaluaciones con fines de selección de personas para ocupar roles de responsabilidad en la conducción de distintos sectores del sistema educativo.

La evaluación del desempeño docente es una de estas formas de evaluación, que tienen lugar cotidianamente en los sistemas educativos, que en general se llevan adelante de manera muy precaria y asistemática, pero que, bien realizadas, tendrían un enorme potencial de cambio. El presente capítulo está destinado a analizar este tema.

La mejora de la docencia es la política educativa más importante

Uno de los aportes de las evaluaciones a gran escala es que ofrecen un panorama comparado de resultados de la acción educativa, en el marco del cual es posible efectuar estudios de profundización, que permitan entender cómo y por qué ciertos países logran que todos sus estudiantes

aprendan y que las desigualdades internas se minimicen. El SERCE puso de manifiesto, en América Latina, los muy buenos resultados de la educación en Cuba, a partir de lo cual algunos investigadores como Carnoy y Marshall (2005) desarrollaron trabajos comparativos que permiten comprender mejor qué es lo que explica los resultados[1].

Uno de los estudios más recientes y enriquecedores es el conocido como "Informe McKinsey", un trabajo de investigación realizado entre marzo de 2006 y mayo de 2007 con el fin de comprender las claves de política educativa que explican que ciertos países, muy diferentes entre sí en términos sociales y culturales, como Finlandia, Singapur, Australia, Corea del Sur y Canadá, logren resultados destacados en términos de las capacidades y conocimientos de sus estudiantes (Barber y Mourshed, 2007).

Después de analizar una gran cantidad de informes de investigación, de entrevistar a más de cien expertos y responsables de políticas educativas, y de visitar escuelas en 25 países, el informe llega a conclusiones muy sencillas y muy contundentes a la vez, tanto en su contenido como en la forma de expresarlas. De acuerdo con el Informe McKinsey, hay tres principios clave detrás de los sistemas educativos con muy buenos desempeños:

1. "La calidad de un sistema educativo no puede ser superior a la calidad de sus docentes".

2. "El único camino para mejorar los aprendizajes es mejorar la enseñanza".

3. "Un buen desempeño como país requiere que todos los estudiantes aprendan".

El estudio muestra, con abundantes referencias, que un buen docente hace una gran diferencia en el aprendizaje de sus estudiantes. Un mal docente, en cambio, en especial en los primeros grados de la escolaridad, puede generar daños irreversibles. Los países que logran los mejores resultados en las evaluaciones internacionales se caracterizan por:

- seleccionar a los candidatos a la formación docente en el tercio superior de los egresados de la educación secundaria;
- ofrecer buenos salarios iniciales, como forma de hacer de la docencia una profesión atractiva;
- ofrecer a los docentes múltiples oportunidades de mejora durante su carrera profesional, con énfasis en la mejora de las prácticas de enseñanza;
- establecer altas expectativas de desempeño para todos los estudiantes;
- establecer mecanismos para identificar tempranamente a los estudiantes con dificultades para alcanzar dichas expectativas y ofrecerles enseñanza adicional.

Como resulta obvio, buenos procedimientos de evaluación son indispensables para la correcta implementación de estas líneas de acción. Hay procesos de evaluación involucrados en:

[1] En general, han resultado mucho más enriquecedores los estudios de carácter cualitativo, realizados en el marco de una evaluación a gran escala, que los llamados "estudios de factores asociados", que solo analizan el comportamiento estadístico de variables medidas en cuestionarios administrados junto con las pruebas.

- la selección de los candidatos que van a cursar carreras docentes (para evaluar no solamente los conocimientos y competencias de tipo cognitivo, sino también las capacidades emocionales y la motivación para la docencia);
- la identificación de quienes no son aptos para la docencia durante los primeros años de ejercicio de la misma (este aspecto es clave dado que, como afirma el informe, una mala decisión puede tener como consecuencia cuarenta años de mala enseñanza);
- la identificación de los profesionales más competentes, con el fin de ofrecerles incentivos que los retengan en la docencia, así como nuevos roles de tutoría y orientación a otros docentes, como modo de difundir su conocimiento práctico al interior de la profesión;
- la creación de mecanismos de evaluación formativa que ofrezcan oportunidades para revisar la propia práctica y para recibir supervisión y orientación (una buena formación inicial es condición necesaria pero no suficiente para tener buenos docentes, es necesario además que estos puedan aprender a partir de la práctica reflexiva y compartida en torno a la enseñanza);
- identificar tempranamente a los estudiantes que no están logrando aprender, así como el tipo de dificultades específicas, como modo de poder intervenir oportunamente (este último aspecto excede el tema de este capítulo, que está referido a la evaluación del desempeño docente).

LA DOCENCIA: UNA PROFESIÓN REÑIDA CON LA EVALUACIÓN

Sin embargo, probablemente una de las razones importantes –entre otras muchas– por las que el cambio es tan difícil en la educación radica en que el rechazo a la evaluación parece formar parte de la cultura docente.

En la mayoría de las actividades humanas la evaluación, entendida como valoración y reflexión en torno a la propia actividad y a los modos de mejorarla, juega un papel central. Los individuos aprenden y mejoran a partir de la evaluación. Piénsese en "oficios" tan dispares como el de un carpintero, un músico y un escritor. En todos estos casos, es natural e inevitable recibir "devoluciones", tanto de los destinatarios del trabajo como de otros colegas con mayor experiencia o de los iguales. Estas devoluciones constituyen un elemento importante y utilizado para mejorar los desempeños y productos del trabajo.

Sin embargo, hay otras actividades que están menos sujetas a "la mirada de los otros" como, por ejemplo, la atención psicológica, la enseñanza y la atención médica. Se trata de actividades que no tienen un producto visible –como un mueble, una pieza musical o un libro– y que involucran la relación personal con otros individuos, relación que suele desarrollarse en espacios privados, como el aula o el consultorio.

En estos casos, el resultado del trabajo es de carácter subjetivo, se produce en la interioridad de las personas[2] y, por tanto, no es directamente visible. Por otro lado, existe una marcada jerarquía de saberes[3] entre el profesional y el estudiante, paciente o cliente, que limita las posibilidades de estos

[2] También la pieza musical o libro tienen un efecto subjetivo, pero existe un producto previo a la vista de todos. No ocurre lo mismo con el trabajo de los psicólogos, médicos y docentes.

[3] También los carpinteros, músicos y escritores poseen un saber especializado pero, debido al carácter estético o comercial de sus productos, por definición todos los destinatarios de su trabajo pueden juzgar-

de formular "devoluciones" –jerarquía esta que, normalmente, la mayoría de los profesionales que ostentan el saber se encargan de explicitar y remarcar.

Finalmente, en estos tres ámbitos de actividad, la relación entre el profesional y el educando, paciente o cliente se produce en espacios físicos a los que normalmente no tienen acceso otras personas, ni siquiera otros colegas, salvo que esto sea explícitamente acordado previamente. Pero no se trata de actividades que puedan ser observadas espontáneamente por otros individuos que no sean los involucrados en la relación de trabajo.

Esto hace que sea particularmente importante, en este tipo de actividades profesionales, generar procedimientos y dispositivos específicos para producir "devoluciones" que permitan valorar y reflexionar sobre la práctica, esto es, evaluaciones.

Pero hay además un elemento diferencial entre la educación y los otros tipos de actividades mencionados, que hacen que la necesidad de evaluación sea más imperiosa: la educación es un derecho de los ciudadanos, que el Estado debe asegurar y que, normalmente, incluye como consecuencia un currículo, es decir, una definición de lo que todos los estudiantes tienen derecho a aprender. Esto no ocurre en las actividades del primer tipo. Si al público, a la crítica o a los colegas no les gusta la obra que produjo el músico, este puede simplemente decir "pues entonces no la escuchen, a mí sí me gusta, y yo compongo para expresarme, no para agradarles". No siempre ocurre así, porque el artista necesita tener un público, pero puede ocurrir y es legítimo.

Lamentablemente, el modo de razonar recién descrito tiene fuerte presencia en los ámbitos educativos. A veces de manera explícita y otras de modo implícito, los docentes dicen o piensan "pues yo enseño esto así, si los estudiantes aprenden o no, el problema es de ellos". Esta actitud no se condice con el rol social y profesional que corresponde a los educadores, en tanto responsables de la puesta en práctica del derecho a la educación.

En defensa de los docentes debe decirse que el aprendizaje no es resultado exclusivamente de lo que ellos hacen, sino también, y principalmente, del esfuerzo del estudiante. Pero este hecho no exime a los docentes de realizar todo lo que sea necesario y posible por mejorar sus modos de enseñar. Debería ser un imperativo ético de la profesión poner más atención en lo que los estudiantes necesitan aprender y en las dificultades que tienen, que en lo que al docente le gusta y le interesa enseñar.

La profesión inevaluable

La aversión a la evaluación está fuertemente arraigada en la cultura de la profesión docente. En primer lugar, porque evaluar a los estudiantes es una de las actividades más ingratas y aburridas en la profesión –no porque necesariamente deba ser así, sino porque se suele hacer mal–. En segundo lugar, porque prevalece la idea de que, en definitiva, la educación no puede ser evaluada, en virtud de su complejidad y subjetividad. No puede ser evaluado ni el aprendizaje de los estudiantes ni el trabajo del docente. Para algunos docentes es una cuestión de dificultad práctica. Para otros, una cuestión de principios, una postura ideológica contraria a cualquier evaluación, en tanto implica un ejercicio de poder.

---

lo de acuerdo con sus propios criterios. No ocurre lo mismo en la enseñanza.

Esto último se traduce en la defensa acérrima de la libertad de cátedra (el docente enseña lo que le parece, independientemente de que sea o no relevante para los estudiantes); en una oposición cerrada a cualquier mirada externa sobre el propio trabajo; e, incluso, en la negativa a "coordinar" con otros profesores, "porque eso me limita en lo que yo quiera enseñar"[4]. Existe en muchos docentes una concepción platónica de la enseñanza[5]: educar consiste en recorrer las aulas hablando de lo que a uno le parece apropiado, "seguido" por los estudiantes que puedan y quieran entenderlo; los demás pueden seguir por otro camino. Wiggins (1998) define la docencia como una profesión "egocéntrica", en el sentido piagetiano del término: los educadores solemos tener dificultad para salir de nuestro propio punto de vista, tenemos dificultad para escuchar y comprender lo que ocurre con nuestros estudiantes, tenemos dificultad para darnos cuenta de cuándo nuestra enseñanza no es clara o adecuada. No logramos imaginar cómo es posible que temas que son tan *obvios e importantes para nosotros* no lo sean para los estudiantes. Tendemos a escuchar lo que queremos y, como los estudiantes necesitan demostrarnos que entienden y que saben, se genera una suerte de complicidad: nosotros hacemos como que les enseñamos, ellos hacen como que aprenden. Pero no comprendemos qué es lo que realmente está ocurriendo.

Parte de esa dificultad para des-centrarse, para salir de los propios puntos de vista e ir al encuentro de los estudiantes, está relacionada con la preocupación central por "cubrir el programa", por "dar" los temas. "Necesito 'dar' estos temas". "Necesito 'cumplir' el programa". Si los estudiantes aprenden o no, no es mi problema.

El otro componente central del problema es el aislamiento que caracteriza el ejercicio de la docencia, particularmente en la educación media. Cada profesor enseña su curso como mejor le parece y puede. Cualquier revisión formal del trabajo docente es percibida como una amenaza o como una intrusión, dado que "nadie puede entender lo que ocurre en mi clase como yo" (Wiggins, 1998, p. 297). Son escasas las instancias de trabajo conjunto, que impliquen una mirada común y compartida sobre lo que se quiere que los estudiantes aprendan –más allá de las declaraciones generales de principios–. No es casualidad que este hecho, el poseer una visión compartida, explícita y comunicada con claridad a los estudiantes con relación a lo que se espera que aprendan, sea uno de los factores que aparecen como relevantes para el logro educativo en las investigaciones sobre el tema.

Al respecto, Wiggins (1998, p. 302) se pregunta por qué aceptamos que los desempeños de los estudiantes permanezcan guardados detrás de una puerta siempre cerrada. ¿Por qué no es la norma que todas las tareas, pruebas y muestras del trabajo de los estudiantes estén a la vista, con los comentarios del docente, como forma de promover la evaluación entre pares y responder a las demandas de padres y supervisores?

Este aislamiento y encierro, esta falta de apertura a "la mirada de los otros", que caracteriza a la docencia, es una de las razones más poderosas que impiden el cambio y la mejora en la enseñanza.

Si los docentes como cuerpo profesional no asumen la responsabilidad de evaluar y adecuar de forma permanente su trabajo en función de lo que ocurre con los estudiantes, simplemente

[4] Palabras textuales de un docente uruguayo en una reunión institucional de coordinación en una secundaria pública.

[5] La mayoría de las veces de un modo implícito, no dicho, a veces inconsciente.

cada uno estará satisfaciendo su visión personal de cómo deben hacerse las cosas. Pero la misión profesional consiste en generar capacidades específicas en los estudiantes: madurez intelectual y emocional, capacidad de comprensión de fenómenos complejos, capacidad de reflexión, conocimientos (Wiggins, 1988, p. 300). Hacerlo bien requiere abrirse al escrutinio de los demás, de los pares profesionales en primer lugar, pero también del resto de los interesados en la tarea educativa (familias, autoridades, sociedad civil, entre otros).

La resistencia a la evaluación

La evaluación externa y la evaluación del desempeño docente son particularmente resistidas y rechazadas por diversas razones, algunas de ellas legítimas, como lo es el hecho de que muchos sistemas de evaluación están tan mal diseñados e implementados, que generan más problemas que aportes.

Tenti (2003) destaca la paradoja de que los docentes son probablemente los profesionales que más evalúan (a sus estudiantes) y, al mismo tiempo, los que más rechazan toda forma de evaluación externa sobre su propio trabajo. En el fondo, propone Tenti, se trata de la resistencia a ser observados por una sociedad que duda de la calidad del trabajo que los docentes están realizando.

> "Durante los últimos años, por una serie de circunstancias complejas [...], (los docentes y las instituciones educativas) ya no son solo sujetos activos de la evaluación, sino que ahora son también objeto de evaluación [...]. ¿Qué es lo que justifica que luego de tantas evaluaciones a las que son sometidos los alumnos desde que entran hasta que salen de determinado nivel educativo sean nuevamente objeto de examen, esta vez por instancias extraescolares [...] generalmente ubicadas dentro del ámbito del servicio público? A primera vista es tentador decir que el sistema educativo es examinado porque existe una sospecha [...]: la 'sociedad' desconfía de lo que hace y produce el sistema educativo (incluso sus propios exámenes) [...]. Todo parece indicar que sobre el sistema educativo planea una sospecha que puede expresarse del siguiente modo: no existe necesariamente una correspondencia entre el conocimiento y las competencias efectivamente en los alumnos y los certificados, diplomas o títulos efectivamente distribuidos por las instituciones escolares [...]. En este contexto, no debería extrañar que los propios agentes escolares se hayan sentido un tanto incómodos frente a este tipo de políticas que ponen un manto de duda sobre lo que ellos hacen y producen cotidianamente en sus aulas" (Tenti, 2003, pp. 166-168).

En la misma obra, Navarro (2003) afirma que en realidad el común de los docentes está más dispuesto a que la profesión sea evaluada de lo que suele aparecer públicamente a través de las posturas sindicales. Según Navarro, tres variables inciden sobre una actitud positiva hacia la evaluación del desempeño docente: 1) trabajar en condiciones favorables; 2) el clima escolar y el sentido de eficacia y logro en la enseñanza; y 3) la edad (hay una mayor apertura a la evaluación entre los más jóvenes). Sin embargo, aun cuando exista una predisposición positiva hacia la posibilidad teórica de la evaluación, suele haber rechazo a las experiencias concretas que se han implementado en la región en la última década, como consecuencia de las debilidades de los propios sistemas implementados. "Ha sido muy difícil poner a funcionar sistemas que representen una expresión competente y sostenible en el tiempo [...]; poner en práctica el principio de evaluación es mucho más complejo y difícil que aceptar el principio en teoría" (Navarro, 2003, pp. 154-159).

> "Las experiencias revisadas en la sección precedente, sin embargo, sugieren hacerse por lo menos la pregunta de si el fracaso de los intentos por introducir evaluaciones del desempeño docente pueden atribuirse a las presiones sindicales y a la oposición de los docentes, o si estas últimas son más bien

> el producto de intentos fracasados y poco convincentes desde el punto de vista técnico y gerencial de implantar el principio de evaluación" (Navarro, 2003, p. 159).

CONFUSIONES CONCEPTUALES

Buena parte del problema de la evaluación docente tiene que ver con confusiones conceptuales acerca de los fines de la evaluación, así como en torno a la naturaleza del trabajo docente y al tipo de incentivos que pueden funcionar en la docencia.

Los fines de la evaluación

La primera confusión se deriva de la omisión en distinguir entre evaluaciones de tipo formativo, dirigidas a propiciar el aprendizaje de los docentes y la mejora de sus prácticas, de las evaluaciones con consecuencias, como las destinadas a calificar al docente, a entregarle un reconocimiento salarial o a clasificar a los docentes en función de sus conocimientos y competencias profesionales. Normalmente se habla de evaluación docente sin distinguir entre unas y otras (Ravela, 2006).

En la práctica habitual de la supervisión en muchos de nuestros países, ambas finalidades se encuentran presentes en el mismo acto de evaluación, lo cual promueve actitudes contradictorias en quien es evaluado. Si la evaluación tiene como finalidad exclusiva ayudarme a identificar mis dificultades y encontrar caminos para superarlas, mi actitud será mostrar esas debilidades. En cambio, si la evaluación tiene como finalidad calificarme, legítimamente mi actitud será la contraria: intentaré mostrar la mejor faceta de mi trabajo profesional y evitaré ofrecer flancos débiles (Ravela, 2008).

Cuando ambas finalidades están presentes en el mismo acto de evaluación, predomina el temor a la calificación y la necesidad obvia de ser bien calificado. Como además la evaluación se realiza con un instrumental totalmente rudimentario desde el punto de vista técnico, la evaluación no cumple adecuadamente con ninguna de las dos finalidades y se torna un ritual burocrático.

De lo anterior se deriva una primera línea de trabajo para la evaluación docente: es necesario diferenciar los dispositivos de evaluación dirigidos a promover el aprendizaje profesional y la mejora de la enseñanza, de los dispositivos de evaluación dirigidos a calificar o categorizar de alguna forma a los docentes.

Los incentivos en la docencia

La segunda confusión se deriva de trasladar a la docencia el tipo de incentivos que funcionan en otros ámbitos de actividad y de trasladar la lógica de la empresa al campo educativo. Todas las personas, incluidos los docentes, aspiran a recibir por su trabajo un salario que les permita vivir dignamente. Una parte importante de las personas –aunque no necesariamente todas– aspiran a que, si realizan un esfuerzo especial y destacado en su trabajo, ello se vea reconocido en su remuneración. Pero solo algunas personas y una parte de las organizaciones colectivas tienen como objetivo principal en la vida hacer mucho dinero.

Si bien la mayoría de las empresas tienen como finalidad principal generar ganancias en dinero, existen muchos otros tipos de organizaciones que se mueven por otros fines y, por tanto, por otros incentivos que el dinero. Entre ellas, la gran mayoría de las organizaciones educativas. Y para la gran mayoría de las personas que dedican su vida a la educación existen múltiples

fuentes de satisfacción –y, por tanto, de incentivos–: el trabajo bien realizado, el entusiasmo y el reconocimiento de los estudiantes, la calidad de las relaciones interpersonales, la profundización en los conocimientos de la disciplina que se enseña, el éxito de los estudiantes en su aprendizaje, el compartir la preparación de su trabajo con colegas, el dar a conocer su trabajo, el participar de una comunidad intelectualmente creativa y desafiante, entre otros, además del salario.

En este sentido, la evaluación de los docentes puede operar como generador de incentivos de una manera mucho más amplia que la de otorgar más dinero a quienes trabajan mejor. Puede ayudar a realizar mejor el trabajo y, por tanto, a crear toda la gama de incentivos que del trabajo bien hecho y del logro de los estudiantes se deriva.

Pago por resultados y pago por mérito

Otra confusión habitual, relacionada con el denominado "pago por mérito", es la relativa a la vinculación de la remuneración a los resultados de los estudiantes en pruebas estandarizadas, por un lado, o al buen desempeño de la labor profesional, por otro. Obviamente hay cierta relación entre buen desempeño y resultados, pero esta relación no es perfecta ni mucho menos.

El aprendizaje de los estudiantes depende del buen desempeño del docente, pero depende también de otros múltiples factores: la nutrición, el apoyo familiar, la historia académica y los maestros que el estudiante tuvo anteriormente y, sobre todo a partir del nivel medio, la voluntad de aprender del propio estudiante, y el tiempo y esfuerzo que está dispuesto a dedicarle al aprendizaje (Ravela, 2006).

Si bien es indiscutible que un buen docente hace una diferencia en lo que aprenden los estudiantes, también lo es el hecho de que un mismo muy buen docente obtiene resultados muy diversos según la motivación de los estudiantes con los que trabaja, su historia escolar previa y las peculiares características del grupo que tiene delante.

El pago por resultados es un camino plagado de problemas. A los ya mencionados es necesario agregar que es de difícil implementación –requiere evaluaciones censales permanentes en todas las disciplinas, además de controles de múltiples variables de contexto–, que desconoce el hecho de que las pruebas estandarizadas evalúan solo una parte de los objetivos educativos, que se presta para promover la reducción del currículo y a generar múltiples conductas reñidas con la ética.

El pago por desempeño tiene más sentido. En Cuba, por ejemplo, cada docente es evaluado todos los años, como consecuencia de lo cual es calificado como "muy bueno", "bueno", "regular" o "insuficiente". Los docentes tienen un salario base, un componente salarial por antigüedad y un componente salarial por desempeño, que tiene tres categorías. Los docentes que son calificados con "muy buenos", al año siguiente suben un escalón en el salario por desempeño, al tiempo que los calificados con "regular" bajan un escalón. Los docentes calificados con "insuficiente" (alrededor del 3% cada año) son retirados del aula y deben realizar una nueva etapa de formación durante un año[6].

[6] Más detalles pueden encontrarse en los materiales aportados por el director del Instituto Central de Ciencias Pedagógicas de Cuba, Héctor Valdés, con ocasión de un Seminario sobre Evaluación del Desempeño Docente, dictado en mayo de 2008 en el Instituto de Evaluación Educativa de la Facultad de Ciencias Humanas de la Universidad Católica del Uruguay. Disponible en la siguiente dirección de Internet:

Sin embargo, un sistema de pago por desempeño no debería limitarse a otorgar un reconocimiento salarial, sino que debería formar parte de un sistema que identifique, retribuya y, sobre todo, capitalice la capacidad profesional de los mejores docentes. Esta propuesta será desarrollada con mayor amplitud en los apartados que siguen.

El desafío principal para la evaluación en este campo no es tanto propiciar la distribución de premios en dinero, como ayudar a identificar a los mejores docentes para otorgarles funciones de mayor responsabilidad en la mejora de la enseñanza –y una remuneración acorde a ello–, así como también identificar a los docentes con dificultades para realizar su labor de manera apropiada para ayudarles a mejorar su trabajo. Uno de los errores de muchas propuestas de pago por desempeño radica en ignorar que, si bien eso puede servir de reconocimiento para quienes hacen bien su trabajo –lo cual es en sí mismo valioso–, no resuelve el problema de quienes no tienen suficiente conocimiento de la disciplina que deben enseñar o no saben cómo enseñarla.

> "Si bien algunas reformas incrementaron los mecanismos de rendición de cuentas o introdujeron incentivos basados en el desempeño para mejorar la motivación, lo hicieron sin brindar a los docentes oportunidades para darse cuenta de sus debilidades o conocer mejores prácticas" (Barber y Mourshed, 2007, p. 27).

Con lo cual volvemos al inicio de este apartado. Son necesarios dos tipos de dispositivos de evaluación docente: unos que contribuyan a la creación de nuevas capacidades y otros que establezcan mecanismos para identificar y reconocer a los buenos docentes.

Esta tensión entre evaluación formativa y evaluación con consecuencias se deriva de una contradicción inherente a los seres humanos. Todos tenemos una tendencia natural a crecer y a ser mejores, pero simultáneamente coexiste en nuestro interior una tendencia a la rutina, a quedarnos en lo que tenemos, a perder el sentido de lo que hacemos. Por eso, a la vez que motivaciones internas, necesitamos incentivos y exigencias externas.

Lo que resta de este capítulo está destinado a la reflexión sobre estos dos conjuntos de dispositivos de evaluación del desempeño docente con finalidades diferentes. Uno tiene que ver con la motivación interna, con el aprendizaje y con la mejora de la práctica. El otro tiene que ver con la exigencia y el reconocimiento externos, con la acreditación y con la carrera docente.

EVALUACIÓN, INCENTIVOS Y CARRERA DOCENTE

Todo sistema educativo de gran escala necesita contar con un buen sistema de evaluación docente, dirigido a establecer categorías profesionales en función de los conocimientos, competencias y responsabilidad en el cumplimento de la tarea. Esto es de vital importancia por dos motivos fundamentales.

En primer lugar, para establecer reconocimientos formales para los docentes que mejor realizan su función. En la actualidad, la mayoría de los sistemas educativos en nuestra región otorgan el mismo tratamiento funcional y salarial a todos los docentes, independientemente de la calidad y compromiso profesional con que realicen su tarea. Las diferencias tienen que ver con la antigüedad o con el tipo de establecimiento en que el docente se desempeña, pero no con la calidad de su trabajo.

---

http://www.ucu.edu.uy/Home/FacultadesyÁreas/CienciasHumanas/Institutos/EvaluaciónEducativa/Actividades/Formación/tabid/1262/Default.aspx.

Esto, en el largo plazo y en un contexto en que muchos docentes adolecen de serias carencias formativas, en tanto otros incumplen en forma reiterada sus obligaciones básicas –asistir a sus clases, prepararlas adecuadamente, preocuparse por el aprendizaje de todos los estudiantes–, desestimula a los mejores docentes y conduce a una cultura de autocomplacencia, ausencia de exigencias y falta de compromiso que además se transmite a los estudiantes.

Esta situación puede ilustrarse recurriendo a los versos de Enrique Santos Discépolo, en su tango *Cambalache*:

Hoy resulta que es lo mismo
ser derecho que traidor...
Ignorante, sabio, chorro,
generoso, estafador...

Todo es igual. Nada es mejor.
Lo mismo un burro que un gran profesor.

No hay aplazados, ni escalafón,
los inmorales nos han igualado...

Con estos versos podría caracterizarse la política de recursos humanos de la mayoría de los sistemas educativos en nuestra región.

En segundo lugar, un sistema de evaluación de esta naturaleza es vital para que el sistema educativo pueda identificar a los buenos docentes con el fin de encomendarles nuevas tareas. En la actualidad, el sistema educativo es "ciego" con relación a sus propios recursos humanos. Carece de mecanismos para identificar, promover y aprovechar la capacidad de sus mejores docentes, simplemente porque no sabe quiénes son ni dónde están. El talento y el valor personal y profesional que poseen estos docentes es desperdiciado por el sistema. Solo lo conocen y aprovechan algunos de quienes se encuentran en sus entornos inmediatos, pero el sistema como tal no lo hace.

Para articular estos dos propósitos sería necesario un sistema de carrera docente de tipo "escalar" (Morduchowicz, 2002), que se caracteriza por establecer categorías profesionales que suponen niveles crecientes de competencia profesional, entre las que se transita a través de mecanismos formales de evaluación del desempeño. Cada categoría supone una mejora salarial con relación a la anterior, así como la asunción de nuevos roles y responsabilidades.

El Cuadro 1 ilustra un sistema de este tipo (Ravela, 2008). Los docentes participarían del mismo de forma voluntaria y en los momentos en que lo deseen, tantas veces como lo deseen. Una evaluación sistemática, a través de diversidad de instrumentos, daría lugar a la ubicación del docente en una categoría profesional, en una escala de cuatro posiciones.

A medida que el docente avanza en esta escala, mejoraría su remuneración y el reconocimiento profesional pero, sobre todo y más importante, su carga de horas en aula se reduciría (sin eliminarse nunca del todo) para asumir nuevas funciones: acompañamiento a futuros docentes en proceso de formación, tutoría a estudiantes con dificultades, inducción de nuevos docentes, elaboración de materiales curriculares, coordinación de equipos de docentes, evaluación de otros docentes.

Esta lógica de carrera docente tiene como virtud principal, más allá del reconocimiento a los buenos docentes, que generaría un círculo virtuoso en el sistema educativo, en la medida que se crean nuevas funciones –que son indispensables– y que los roles de orientación y modelamiento son ocupados por los mejores docentes –cosa que nuestros sistemas de selección actuales no garantizan en absoluto, con lo cual muchas funciones, como la del supervisor, están fuertemente devaluadas.

Cuadro 1. Una propuesta de sistema de evaluación y de carrera docente articulados

Nivel o posición	Funciones	Requisitos de acceso
Etapa inicial	• Enseñanza en carácter interino.	• Título docente.
Etapa profesional	• Enseñanza con carácter de titular. • Orientación o tutoría de estudiantes.	• Cinco años en el nivel anterior. • Informes favorables de desempeño de sus directores. • Pruebas de conocimientos disciplinares y de didáctica.
Profesional avanzado	• Enseñanza. • Tutor de práctica docente. • Seguimiento y orientación a nuevos docentes. • Elaboración de materiales curriculares y guías para la enseñanza.	• Cinco años en el nivel anterior. • Informes favorables de desempeño de sus directores. • Pruebas de didáctica. • Presentación de carpeta de materiales de enseñanza empleados en sus clases. • Curso de capacitación para orientación de docentes.
Profesional experto	• Enseñanza. • Coordinador de departamento o asesor pedagógico. • Formador de docentes en servicio. • Evaluador de pares.	• Cinco años en el nivel anterior. • Informes favorables de desempeño de sus directores. • Pruebas de conocimientos avanzados en su disciplina. • Presentación de carpeta que recoja su trabajo como orientador de docentes. • Presentación de carpeta de materiales curriculares elaborados. • Curso de capacitación para evaluación de docentes.

La construcción de un sistema como el propuesto tiene varios requisitos importantes. En primer lugar, es preciso construir un marco conceptual apropiado y de amplia aceptación acerca de qué es buen desempeño docente. Ello requiere el involucramiento de la profesión docente. En segundo término, la legitimidad del sistema se apoya sobre el diseño de un conjunto de instrumentos y procedimientos de alta calidad técnica, que den lugar a procesos de evaluación rigurosos, válidos, justos y reconocidos como legítimos por el cuerpo docente. En tercer lugar, es necesario cons-

tituir progresivamente un cuerpo de evaluadores cuya idoneidad sea ampliamente reconocida. Con el paso del tiempo, el propio sistema permitirá identificar a docentes expertos para que actúen como evaluadores.

Entre los instrumentos centrales para los procesos de evaluación del desempeño docente que un sistema así involucraría cabe destacar los siguientes:

- Informes del desempeño del docente en los centros educativos en los que trabaja, que incluyan aspectos tales como el cumplimiento de las obligaciones funcionales, la participación en la comunidad educativa, el vínculo con los estudiantes y el compromiso con los proyectos institucionales.
- Pruebas de dominio de los conocimientos de las disciplinas que el docente enseña.
- Pruebas de didáctica de dichas disciplinas.
- Carpetas o portafolios con muestras de diversas facetas de su trabajo.
- Observación del docente en el aula y entrevista en detalle por parte de evaluadores pares o supervisores, especialmente entrenados para la tarea.

Un sistema como el propuesto tendría la potencialidad de provocar, a medio plazo, una dinámica positiva en varias dimensiones del sistema educativo:

- Los mejores profesores estarían orientando a los nuevos docentes, supervisando a sus colegas, evaluando y seleccionando a sus futuros reemplazos.
- El sistema comenzaría a otorgar reconocimiento simbólico y económico a los buenos docentes.
- Permitiría vincular mejoras salariales importantes con capacidad y compromiso profesional reconocidos y con el desempeño de nuevas funciones.
- Abriría nuevas oportunidades profesionales que servirían para retener buenos docentes en el sistema.
- Enriquecería al sistema educativo con nuevas funciones, indispensables e inexistentes, que serían desempeñadas por los docentes más competentes.
- Permitiría al sistema aprovechar y crear un efecto multiplicativo a la capacidad de sus mejores docentes.

EVALUACIÓN, CONOCIMIENTO PRÁCTICO Y MEJORA DE LA ENSEÑANZA

La propuesta incluida en el apartado anterior está basada en un esquema de evaluación con consecuencias, articulado con un sistema de carrera docente que involucra los aspectos salariales y de funciones docentes.

En paralelo con el anterior, sería necesario diseñar un conjunto de dispositivos dirigidos a "romper" el carácter cerrado, aislado y "egocéntrico" –en sentido piagetiano del término– del ejercicio de la profesión, con el fin de transformarla en una profesión abierta a "la mirada de los otros" y al aprendizaje continuo a partir de la reflexión sobre la práctica.

Abrir realmente el aula a los pares y a supervisores u orientadores pedagógicos requiere de un dispositivo de evaluación sin consecuencias funcionales que puedan constituir amenazas para el evaluado, pues esto, si bien legítimo para otra finalidad, pervierte la finalidad formativa. Si el docente va a ser calificado de algún modo, el carácter formativo de la evaluación se esfuma.

Un antecedente interesante en la región lo constituye El Salvador, que hace algunos años transformó el cuerpo de inspectores en un cuerpo de asesores pedagógicos, con la función exclusiva de brindar orientación pedagógica a los docentes. Simultáneamente se mantuvo un sistema de control administrativo paralelo, a cargo de un cuerpo de personal técnico-administrativo diferente.

En esta línea de pensamiento, el "Informe McKinsey" destaca el foco en la mejora de la práctica en el aula, como una de las características centrales de los sistemas educativos cuyos estudiantes han logrado los mejores desempeños en pruebas internacionales. Estos países se caracterizan por:

- Poner un foco importante en construir habilidades prácticas para la enseñanza desde la formación inicial –por oposición a la formación exclusivamente teórica que caracteriza a muchos sistemas–.
- La constitución de equipos de docentes expertos que apoyan el desarrollo profesional de los docentes en las escuelas, modelan la enseñanza y promueven el aprendizaje continuo.
- Procedimientos apropiados y efectivos para la selección y desarrollo de líderes de enseñanza.
- Construcción de espacios de intercambio de experiencias y aprendizaje mutuo al interior de las escuelas y como parte de la labor cotidiana de los docentes.

> "[...] algunos de los mejores sistemas han encontrado vías para hacer posible que los docentes aprendan unos de otros. En la mayoría de las escuelas los docentes trabajan en forma aislada. En varios de los mejores sistemas educativos, especialmente en Japón y Finlandia, los docentes trabajan en forma conjunta, planifican sus clases juntos, se observan mutuamente en el aula y se ayudan a mejorar unos a otros. Estos sistemas crean una cultura en sus escuelas según la cual la planificación colectiva, la reflexión sobre la enseñanza y el acompañamiento entre pares constituyen la norma y una característica constante en la vida escolar. Esto permite a los docentes aprender en forma continua" (Barber y Mourshed, 2007, p. 28).

Para captar la importancia de estos dispositivos es preciso comprender los diferentes tipos de conocimiento que se ponen en juego en el ejercicio de la docencia. Por cierto, todo docente necesita ser portador de conocimientos sobre la disciplina que va a enseñar. Pero dominar la Matemática o la Filosofía es condición necesaria pero no suficiente para enseñarlas de manera apropiada. Tampoco es suficiente estudiar teóricamente la didáctica de la disciplina, es necesario desarrollar un conjunto de conocimientos prácticos, un tipo de conocimiento que no puede ser codificado ni transmitido a través del lenguaje escrito. Se trata de una forma de conocimiento que las personas llevan incorporado (*in-corpore),* es parte de su personalidad (Tenti, 2001). Esta forma peculiar de conocimiento solo puede ser transmitida de unos a otros a través del contacto personal y del intercambio de experiencias.

Este hecho explica el fracaso de muchos esfuerzos de formación en servicio en los países de la región. En la medida en que los mismos están orientados exclusivamente al plano de los discursos y las ideas, no alcanzan a producir cambios en las prácticas de enseñanza en el aula y, como consecuencia, tampoco en los aprendizajes. En dichos programas, los docentes básicamente aprenden

los discursos y términos pedagógicos de moda y políticamente correctos, pero no desarrollan habilidades prácticas para modificar sus modos de trabajar en el aula.

De allí la importancia de implementar dos tipos de dispositivos de evaluación formativa: la evaluación por parte de expertos y la evaluación a través del intercambio entre pares.

La evaluación formativa por parte de expertos sirve para ofrecer orientación y modelos a los nuevos docentes y a los docentes menos competentes. Muchas destrezas, conocimientos prácticos y actitudes profesionales solo se adquieren en la relación directa con un docente experto. Obviamente, esto requiere de una adecuada identificación de estos expertos, para lo cual un sistema como el descrito en el apartado anterior resulta crucial.

La evaluación colectiva de las prácticas de trabajo en el aula es la otra gran avenida para construir y compartir el conocimiento práctico que buena parte de los docentes genera a diario en el aula, pero que no trasciende más allá de la puerta del salón de clases. Pero no basta con crear espacios de trabajo colectivo, es preciso que estos tengan contenido y liderazgo. De lo contrario se transforman en espacios de conversación, que tampoco afectan a las prácticas de enseñanza. En varios países de la región, entre ellos Uruguay, se han creado espacios de coordinación para los docentes y horas de departamento, pero que no han logrado constituirse en un espacio de mejora de la enseñanza, simplemente porque no se ha logrado que los docentes abran sus aulas a sus colegas y se focalicen en el análisis de las prácticas de enseñanza.

Los instrumentos de evaluación para facilitar esta tarea son muy evidentes (la visita al aula, el registro y análisis posterior de clases en vdeo, la construcción y revisión colectiva de portafolios o carpetas de trabajo, entre los más relevantes), aunque su implementación no es sencilla, tanto por dificultades de tipo práctico (como, por ejemplo, coordinar los horarios de manera que los docentes puedan visitarse mutuamente en sus aulas), como de tipo cultural (el temor a abrir el aula a los colegas).

Wiggins (1998) propone, como primer paso en este camino, transparentar las evaluaciones que hacemos de nuestros estudiantes a los demás colegas. Revisar colectivamente las propuestas, consignas y criterios de evaluación puede ser un punto de partida para iniciar procesos de revisión de la práctica entre pares y la creación de relaciones realmente profesionales. Puede ayudarnos a salir de nuestra –inevitablemente– distorsionada percepción de qué están aprendiendo los estudiantes.

Necesitamos, sugiere Wiggins, aprender a ser receptivos a las pruebas disponibles acerca de la brecha inevitable entre nuestras intenciones educativas y nuestros efectos reales en los estudiantes. No se trata de establecer premios, ni castigos, ni culpas. Simplemente se trata de analizar los datos empíricos disponibles sobre lo que está ocurriendo con mis estudiantes, en mis clases. Aprender a ver lo que ocurre en el aula y con cada uno de los estudiantes es el primer paso para llegar a ser un buen docente. Para esto necesitamos ayuda, necesitamos abrirnos a la mirada de los otros. La evaluación de carácter formativo puede ser una herramienta central para ello.

Finalmente, lo que importa es comprender que, si el desafío contenido en este último apartado no se enfrenta de manera exitosa, difícilmente puedan ocurrir cambios en los modos de enseñar. La enseñanza se torna una rutina más o menos confortable, que permite sobrevivir al día a día con los estudiantes. Y sin cambios en las formas de enseñar, difícilmente se produzcan cambios en el aprendizaje de los estudiantes.

Evaluación de los aprendizajes en el aula: una conceptualización renovada

Juan Manuel Esquivel

INTRODUCCIÓN

Este capítulo pretende principalmente presentar a los educadores una propuesta de evaluación de los aprendizajes congruente con una concepción del aprendizaje fundamentada en el socio-constructivismo, que reúne las teorías cognitiva y sociocultural. Shepard (2006) señala lo siguiente: "si bien hay muchos puntos discutibles no resueltos entre estas dos perspectivas y en el interior de ellas, las considero compatibles". Esta propuesta se aleja del modelo de evaluación de los aprendizajes que ha predominado en las aulas de nuestras escuelas primarias y secundarias durante las últimas décadas. Tal modelo que podría caracterizarse, en términos generales, por el empleo de pruebas llamadas "objetivas" y por un uso frecuente de la evaluación con el fin de calificar a los estudiantes y, en muchas instancias, definir y decidir aspectos importantes en sus vidas.

Considera la propuesta de este capítulo que la enseñanza, el aprendizaje y la evaluación deben constituirse en una unidad indisoluble (Coll, Martín y Onrubia, 2001), de manera tal que la evaluación sirva para explicar tanto los aprendizajes de los estudiantes como la actividad instruccional del maestro. La evaluación entonces servirá para ofrecer información a los estudiantes acerca de los aprendizajes alcanzados, a fin de que puedan asumir la responsabilidad de controlar sus logros. Por otra parte, servirá al docente para obtener información que le permita mejorar su actividad didáctica al planear y ejecutar actividades de enseñanza.

La evaluación, desde la perspectiva señalada en el párrafo anterior, ofrecerá a los actores del hecho educativo información útil y relevante para mejorar cada uno su acción. A los alumnos, para que los conocimientos construidos sean más amplios, significativos y profundos, y a los profesores, para que su actividad didáctico-pedagógica sea más eficaz en cuanto a promover esos mejores aprendizajes. Como muy bien señalan Coll, Martín y Onrubia (2001), "la evaluación de los aprendizajes puede estar o ponerse al servicio de un proceso de toma de *decisiones de naturaleza pedagógica o didáctica*" (p. 4).

En los diferentes apartados de este capítulo se pretende ofrecer a los profesores, en primer lugar, una revisión de algunos conceptos teóricos básicos sobre evaluación de los aprendizajes. Seguidamente, se hace una crítica a la forma como se emplea, en términos generales, la evaluación de los aprendizajes en las aulas de la educación básica y secundaria, y se fundamenta, desde la perspectiva de una concepción socio-constructivista de los aprendizajes, una evaluación congruente con tal concepción. Finalmente, se ofrece una revisión general de los principios de la evaluación del desempeño y de la evaluación auténtica, que incluye información acerca de los principales medios e instrumentos que emplea, y se brinda algunos ejemplos de estos.

REVISIÓN DE CONCEPTOS FUNDAMENTALES

En su conceptualización más amplia, la evaluación de los aprendizajes es un proceso sistemático de acopio de información que permite al profesor emitir un juicio de valor sobre las adquisiciones o aprendizajes que alcanzan sus estudiantes como resultado de su participación en las actividades de enseñanza. La información que se recolecta está condicionada por la existencia de dos elementos fundamentales: 1) ¿Cuáles son los aprendizajes que se pretende que alcancen los estudiantes?, y 2) ¿Cuáles son las muestras observables en las tareas, realizaciones o ejecuciones de los estudiantes que se consideran como pruebas del logro de los aprendizajes? Comúnmente, los aprendizajes se formulan en términos de objetivos, mientras las pruebas se obtienen de la observación de las tareas, realizaciones y ejecuciones de los estudiantes y de la aplicación de escalas y rúbricas a estas.

Por otra parte, es importante analizar las funciones de la evaluación. La evaluación puede darse en diferentes momentos de los procesos de enseñanza y del aprendizaje, y también puede servir para la toma de diferentes clases de decisiones, según el propósito del empleo de la información que genera. De estas dos perspectivas, tiempo y propósito, se originan las tres funciones básicas de la evaluación: diagnóstica o inicial, formativa y sumativa.

La *evaluación diagnóstica* se da al inicio de los procesos de enseñanza y de aprendizaje. Sobre la base del juicio de valor emitido, se toman las decisiones pedagógicas que han servido tradicionalmente para fundamentar el planeamiento didáctico de una unidad o curso en la realidad de los conocimientos previos de los estudiantes. Los resultados de esta evaluación han favorecido además una enseñanza basada en los principios de la educación inclusiva, pues han permitido adaptar la enseñanza a las características y necesidades de los estudiantes (Ainscow, 2001; Arnaiz, 2003, y Esquivel, 2006).

La *evaluación formativa* es apropiada durante el desarrollo de los procesos de enseñanza y de aprendizaje. Permite ofrecer información a los estudiantes y a los profesores sobre los aprendizajes logrados en un momento determinado de estos procesos. Esta conceptualización tradicional se ha visto enriquecida con la propuesta que hace Sadler (1989), quien aportó el modelo más aceptado de la evaluación formativa. Este autor indicó que no es suficiente que los maestros simplemente señalen si las respuestas dadas en una prueba son correctas o incorrectas, o si la tarea ejecutada (o el producto) exhibido refleja aprendizajes significativos. Esta retroalimentación deberá ir necesariamente acompañada de criterios explícitos y claros de desempeño, así como de información a los estudiantes sobre estrategias para facilitar el aprendizaje. Este aporte de Sadler se acerca a la conceptualización aportada por Coll, Martín y Onrubia (2001), según la cual la información que ofrece el juicio de valor deberá ayudar al profesor a tomar decisiones que mejoren sus actividades de enseñanza, y a los estudiantes, a mejorar su aprendizaje. Esta función de regulación de los procesos de enseñanza y de aprendizaje hace que se lo denomine "evaluación reguladora". Por otra parte, según señalan los autores citados arriba, "se ha subrayado su vertiente ´formadora´, es decir, su utilidad para que los alumnos aprendan a regular sus procesos de aprendizaje" (p. 5).

Finalmente, la *evaluación sumativa*, también denominada "acumulativa" o de "resultados", se realiza al terminar un período determinado de los procesos de enseñanza y de aprendizaje. Tiene como propósito principal calificar el grado del desempeño de los estudiantes en el logro de los objetivos curriculares. Esta evaluación se efectúa usualmente al terminar períodos escolares de

diferente duración, como los correspondientes a las unidades de aprendizaje, a las partes del curso señaladas en el calendario escolar, como los trimestrales o los semestrales, al final de los cuales deben entregarse a los padres de familia informes acerca del rendimiento académico estudiantil, y *a todo el curso lectivo*. También se da el caso de sistemas educativos en los cuales los profesores realizan evaluación sumativa en forma programada y sistemática. Con los resultados de la evaluación sumativa el profesor emite un juicio sobre el logro final de los aprendizajes y sobre esta base se asigna al estudiante la calificación numérica o alfabética correspondiente al período evaluado. Si se trata del final de un curso o de un período lectivo, esta calificación permite una certificación con fines de promoción.

Como se ha indicado anteriormente, la evaluación disociada de los procesos de enseñanza y de aprendizaje corresponde a un modelo de evaluación de los aprendizajes fundamentado en la tradición psicométrica, adoptada por la psicología educativa y por una gran mayoría de los evaluadores educativos. En este modelo (clásico o normativo) esencialmente se interpretan los resultados del aprendizaje de un individuo, evaluado mediante pruebas, desde una perspectiva comparativa. En otras palabras, el significado del resultado de un individuo adquiere sentido cuando se compara con los resultados del grupo al que pertenece o al grupo normativo. Sin embargo, existe otro modelo de elaboración y validación de pruebas (modelo referido a criterios) que, por sus fundamentos y procedimientos, difiere sustancialmente del descrito antes, aunque no tiene su solidez teórica y metodológica. En la interpretación de los resultados manifiesta un sentido más pedagógico, pues compara el resultado de un individuo con objetivos o intenciones de logro.

En esta revisión general de conceptos básicos sobre evaluación de los aprendizajes es necesario dedicar un espacio de análisis a algunos otros conceptos importantes. El primero se refiere a la conceptualización errónea del término "medición" como sinónimo de "evaluación". Ya señalamos en un párrafo anterior que la evaluación implica la emisión de un juicio de valor, basado en información relevante. Esta información podría incluir el resultado de la medición del logro de algún objetivo, mediante la definición de indicadores observables en las tareas, exhibiciones o producciones escritas, artísticas o de otra índole de los estudiantes, que se comparan con los objetivos o intenciones.

Al análisis anterior hay que agregar la dimensión comunicativa que tiene la evaluación. La comunicación que emite el profesor de los juicios de valor sobre el resultado del proceso evaluativo siempre tiene un destinatario, sean estos los estudiantes, otros profesores, la colectividad de la institución educativa, los padres y madres y el mismo profesor, y también tiene un propósito. Sin embargo, muchas veces el propósito del profesor puede ser distorsionado por las decisiones que, con respecto al empleo de la información, tomen algunos de sus destinatarios.

LA EVALUACIÓN COMO PARTE INTEGRAL DE LA ENSEÑANZA Y EL APRENDIZAJE

La evaluación disociada de la enseñanza y el aprendizaje

Tradicionalmente, el maestro ha considerado la evaluación como un elemento importante de su quehacer de aula, pero la ha aplicado al final de una unidad o curso para comprobar mediante ella si los alumnos han logrado los aprendizajes esperados. Esa concepción disocia la evaluación de la enseñanza y del aprendizaje y la convierte en un ejercicio que influye muy poco en la modificación de su práctica de la enseñanza y en el mejoramiento de los aprendizajes de los estudiantes.

Como consecuencia de esta realidad, la evaluación, para la mayoría de los maestros, tiene un único propósito: justificar las calificaciones de sus estudiantes. Para estos, la evaluación se constituye en el filtro determinante del éxito o fracaso escolar, muchas veces lejos de la determinación de los verdaderos aprendizajes adquiridos.

Por otra parte, los maestros han recibido una formación en evaluación fundamentada teóricamente en el modelo psicométrico y con énfasis en la elaboración de pruebas según los dictados de la elaboración de los tests, en un modelaje de los tipos de preguntas o ítems que generalmente se emplean en las pruebas estandarizadas. La conceptualización de la evaluación como sinónimo de medición, que generalmente hacen los maestros, es una consecuencia de la importancia que tiene el modelo psicométrico en la formación inicial del educador. Esta conceptualización se refuerza con la presencia, en los cursos de evaluación de los aprendizajes, de una unidad de estadística descriptiva, que deriva en los análisis de calidad de los ítems y en el análisis de confiabilidad y validez dependientes de la comprensión de la estadística y de las premisas básicas del modelo psicométrico: la curva normal y la necesidad de una alta variabilidad de los puntajes. Son muy pocos los maestros que se cuestionan el hecho de que estas dos premisas son esencialmente contrarias a la finalidad última de su papel como docente: facilitar con sus actividades de enseñanza el mayor aprendizaje posible en el mayor número posible de sus estudiantes. Los puntajes de los alumnos que alcancen estos logros no conformarán una curva normal, ni lograrán una dispersión o variabilidad como se esperaría para considerar los parámetros de la prueba y sus ítems como adecuados.

La imitación del evaluador profesional por parte del maestro en la elaboración de preguntas, como se enfatiza en su formación, propugnará la elaboración de preguntas que se caractericen por solicitar al estudiante que seleccione una respuesta verdadera entre varias opciones o que identifique una relación simple, que llene un espacio en blanco, que identifique la veracidad o falsedad de una frase o de preguntas que requieren para su respuesta escrita únicamente la memoria o que son de muy poca elaboración de parte del estudiante. Además se da el caso de que la mayoría de los libros de evaluación de los aprendizajes que se encuentran a la disposición de los maestros[1] hacen énfasis en la elaboración de preguntas como las antes citadas. Todas ellas presentan una característica común: la respuesta se califica como buena o mala, lo cual implica que la calificación del test compuesto por estas se realiza sin la participación del criterio del profesor. De allí que se encuentre que los maestros emplean preguntas de los tipos citados anteriormente, que solicitan del alumno solamente la memorización o la comprensión de relaciones simples. Cada vez es más difícil encontrar en las pruebas confeccionadas por el maestro preguntas de ensayo en las que se solicita al estudiante la construcción de una respuesta.

Rápidamente, el estudiante se convence de que hay unas reglas de juego que conviene seguir, que indican que hay que lograr buenas calificaciones y que, para hacerlo, debe contestar bien las preguntas de una prueba. Esta prueba exige memorizar hechos, relaciones, fórmulas, procedimientos y comprender relaciones simples entre estos contenidos. Como resultado, pocos aprendizajes son significativos. El alumno estudia para responder en la prueba, haciendo énfasis en

[1] He examinado los 18 libros que aparecen más comúnmente en la bibliografía de los cursos de evaluación de los aprendizajes en escuelas formadoras de docentes de Costa Rica, y solamente uno no hace énfasis en la elaboración de preguntas objetivas o caracterizadas por una respuesta que se califica como buena o mala.

la memorización, sin una conciencia clara con respecto a los aprendizajes y su importancia. Lo hace durante un período corto y determinado, cuando se acerca la fecha de cada prueba. Pero esta realidad le enseña la poca importancia que tienen los aprendizajes, que tiene la conciencia de haber alcanzado.

El maestro como protagonista en la rendición de cuentas

Durante los últimos veinte años se ha producido un movimiento muy importante en Latinoamérica, tendente a establecer sistemas nacionales de evaluación del logro de los estudiantes (Rojas y Esquivel, 1998; Ferrer, 2007). Esta tendencia corresponde a una concepción economicista de la educación, que pretende medir la eficiencia de los sistemas educativos mediante la medición del logro de los aprendizajes de los estudiantes. Algunos pocos sistemas nacionales emplean pruebas estandarizadas, cuyos resultados pueden tener consecuencias graves para los estudiantes, pues sirven para decidir la graduación en un determinado nivel de la educación. Las hay también con resultados cuyas consecuencias son menos graves, pues se emplean como información diagnóstica del sistema educativo.

El hecho de que en estos sistemas se empleen pruebas estandarizadas constituye para el maestro una presión para que imite el tipo y la forma de preguntas que se aplican en estos tests. Significa para el maestro una razón más para elaborar pruebas con preguntas de opción múltiple en la evaluación de las intenciones u objetivos correspondientes a aprendizajes que los estudiantes deben alcanzar. Pero las consecuencias de la administración de estas pruebas son muy preocupantes, sobre todo en aquellos casos en que los resultados significan consecuencias graves, pues incentivan la tendencia a enseñar para la prueba.

Las pruebas estandarizadas ofrecen grandes ventajas para la medición de características psicológicas del ser humano y como determinantes generales del rendimiento académico o del logro de los aprendizajes, y pueden cumplir una función válida de rendición de cuentas en un sistema educativo. Sin embargo, tienen poca utilidad en la evaluación de los aprendizajes en el aula y en la determinación de los aprendizajes complejos de los estudiantes, por las limitaciones impuestas por el tipo de pregunta empleada y por el modelo estadístico que lo fundamenta.

Existen datos que demuestran que el empleo de este tipo de pruebas en el aula tiene un efecto perjudicial en la calidad de la enseñanza y el aprendizaje de los estudiantes (Resnick y Resnick, 1992, citado por Shepard, 2006).

La evaluación según una concepción integradora

Como se señaló en la introducción, la evaluación puede ponerse al servicio de la toma de decisiones de naturaleza pedagógica y didáctica. En otras palabras, se confiere una función pedagógica a la evaluación, que se logra por la aplicación de la evaluación formativa, mediante la cual se evalúa continuamente durante todo el proceso de enseñanza y de aprendizaje. Las tareas y medios de la evaluación son una extensión de este proceso, que enfatiza una caracterización cualitativa y dinámica de los aprendizajes. Esta función pedagógico-didáctica es de carácter integrador, en el sentido de que hace posible la integración de la unidad enseñanza, aprendizaje y evaluación. A esta función pedagógica se contrapone una función social de la evaluación, que separa las actividades de la evaluación de los procesos de enseñanza y de aprendizaje. Esta concepción se caracteriza, entre otras razones, por el empleo de la evaluación sumativa; por el énfasis en los re-

sultados cuantitativos y estadísticos, que redunda en la calificación de los estudiantes al final del proceso de enseñanza y de aprendizaje; por el empleo de pruebas escritas de tipo test, hechas por el profesor, pero que imitan las pruebas estandarizadas. Esta función social hace que la evaluación esté disociada de los procesos de enseñanza y de aprendizaje.

En la realidad del proceso educativo que se da en las aulas de nuestras escuelas existen tensiones entre estas dos funciones, la social y la pedagógico-didáctica, pero predomina la función social. Estas tensiones provienen del fuerte impulso que se ha dado a la teoría constructivista para explicar el aprendizaje y dirigir la enseñanza. Ello es evidente en los documentos curriculares y en el discurso de los curriculistas de los Ministerios de Educación, en contraposición con una evaluación que continúa cumpliendo en la mayoría de los casos una función social por la costumbre, por la falta de formación de los curriculistas y los docentes o por reglamentos o normas de evaluación.

Esta situación debe cambiar. Existe la necesidad de que los teóricos del currículo de los ministerios y los profesores comprendan los elementos básicos de una evaluación con función pedagógico-didáctica, fundamentada en la concepción socio-constructivista del aprendizaje, y la apliquen. De esta manera, la evaluación estará relacionada con un modelo cognitivo que no excluye la ayuda social que el enfoque sociocultural aporta con respecto a cómo aprende el estudiante.

Breve conceptualización del aprendizaje socio-constructivista

La concepción cognitiva del aprendizaje constituye una reacción contra la psicología de las diferencias individuales y el conductismo, cuyo interés fundamental era la adquisición de conocimientos, habilidades y destrezas, pocas veces integradas en competencias complejas, que se evidenciaban en conductas observables. En la concepción cognitiva, los aprendizajes son construidos por los aprendices al realizar conexiones entre la nueva información y las estructuras previas de conocimiento. Los esquemas mentales desempeñan la función de organizar el conocimiento. De tal forma, estos pueden ser recuperados y utilizados en nuevas situaciones cuando se necesita. Por otra parte, estos esquemas se aprovechan también en los procesos llamados de "metacognición", que permiten a quienes aprenden monitorear y manejar su propia comprensión y aprendizaje. Otros elementos importantes que enfatizan los cognitivistas son el papel de la comprensión conceptual y la transferencia (el uso del conocimiento en situaciones novedosas), que se hace posible por la adquisición de generalizaciones y el empleo de los esquemas que permiten reconocer similitudes en las distintas tareas que se presentan. Por otra parte, el modelo sociocultural (Vigosky, 1978) enfatiza la naturaleza social del aprendizaje y "la idea de que la competencia y la identidad de quienes aprenden se desarrollan por medio de la participación mediada socialmente en una actividad práctica llena de significado" (Shepard, 2006, p. 17). Esta autora propone una concepción del aprendizaje socio-constructivista, que reúne las teorías sociocultural y cognitiva. Así pues, la primera establece que los procesos sociales respaldan y definen el aprendizaje, y la segunda permite entender los procesos mentales subsecuentes y recurrentes del individuo.

Caracterización de la evaluación con función pedagógico-didáctica

En párrafos anteriores se estableció una caracterización breve de la evaluación para que corresponda a la función pedagógico-didáctica. Las características necesarias para ello se explican con mayor detalle a continuación:

- Los objetivos o intenciones de aprendizaje deben estar claramente definidos y deben ser entendidos de igual forma por el profesor y sus estudiantes. Estas intenciones deberán ser objeto de discusión en la sala de clase y de una debida apropiación por parte de todos los actores: estudiantes y profesor.

- Además deben estar igualmente claros los criterios con que se juzgarán las tareas, demostraciones, exhibiciones y resolución de problemas con que el estudiante demostrará el dominio de los aprendizajes. La elaboración conjunta o la discusión de una propuesta de las hojas o escalas de calificación o rúbricas entre el profesor y los estudiantes constituye una excelente forma de comprensión de esos criterios.

- El desempeño del estudiante deberá ser juzgado en relación con los criterios definidos y no comparando su desempeño con el de sus compañeros (evaluación referida a criterios).

- Los estudiantes deberán participar en la valoración de sus propios aprendizajes y en la valoración del aprendizaje de sus compañeros. En otras palabras, el profesor incluirá oportunidades de autoevaluación y coevaluación en el planeamiento de su proceso de enseñanza.

- La evaluación deberá ser esencialmente formativa y formadora. Estas características implican que el profesor no solo deberá ofrecer numerosas y continuas oportunidades para que el estudiante se enfrente a situaciones problemáticas que le permitan demostrar sus aprendizajes, sino también establecer el clima apropiado para una clara retroalimentación que sea entendible y que ofrezca al estudiante la oportunidad de llenar las carencias y reflexionar sobre su propio aprendizaje. La evaluación formadora requiere no solo que los estudiantes se apropien de los objetivos o intenciones, como se ha explicitado en una caracterización anterior, sino también que logren supervisar su propio mejoramiento y asumir la plena responsabilidad de su aprendizaje.

- Los estudiantes deben realizar procesos mentales más complejos y retadores, pues las actividades de evaluación los obligan a ejercitarse en estos procesos.

- Es importante la comprensión por parte del profesor de las consecuencias motivacionales de la retroalimentación, pues podrían tener consecuencias negativas el empleo de elogios falsos o exagerados y la retroalimentación negativa directa. Facilitar el aprendizaje es la finalidad fundamental de la retroalimentación; por lo tanto, este propósito debe ser bien comprendido por los profesores y los estudiantes.

- La relación estrecha entre enseñanza, aprendizaje y evaluación se ejemplifica con el diseño y la ejecución de actividades de evaluación que se constituyan de la forma más semejante posible en ejemplos de los objetivos reales de aprendizaje. De tal manera, las actividades de enseñanza se convierten en actividades de evaluación.

- Shepard (2006) propone el concepto de "encarnar" como la caracterización completa de la alineación entre las tareas, los proyectos y las preguntas complejas o problemas a que se enfrentan los estudiantes y todos los elementos constitutivos que se han especificado como comprensiones y capacidades que se quiere que logren.

- La evaluación debe ser continua. Las oportunidades de realizar evaluaciones formativas se planifican de tal forma que se presenten durante todo el proceso de enseñanza y de aprendizaje.

- La evaluación se constituye en un recurso para juzgar la docencia y significa una oportunidad de mejoramiento del proceso de enseñanza.
- La calidad de los instrumentos y medios que se utilizan para evaluar es importante. Esta calidad se podría juzgar, desde la perspectiva de las demostraciones de su validez, por lo siguiente:
  1. El alineamiento que presenten con respecto a los objetivos o intenciones y los criterios con los que se juzgan.
  2. La congruencia con las actividades de enseñanza.
  3. Su representatividad entre las posibles situaciones o problemas que ejemplifican las intenciones u objetivos.
- También, la calidad de los instrumentos y medios se podría juzgar por la confiabilidad de los resultados. Metodológicamente podría calcularse la confiabilidad de la aplicación de los criterios para juzgar un logro determinado:
  1. Por parte de todos los estudiantes y el profesor.
  2. En la autoevaluación por parte del estudiante en diferentes ocasiones.
  3. En la coevaluación por parte de los coevaluadores.
- Ante la realidad de que nuestros sistemas educativos privilegian la evaluación sumativa en su acepción acreditativa, porque lo establecen las normas legales o por la costumbre, se debería procurar lo siguiente:
  1. Que tenga un papel menos protagónico en la educación básica. Que en esta se incremente su empleo conforme aumenta el nivel educativo, empezando después del tercer año escolar.
  2. Que se realice de tal manera que se juzgue el desempeño del individuo respecto a los criterios establecidos y no con respecto al desempeño de los otros estudiantes (evaluación criterial en vez de evaluación normativa).
  3. Que los resultados obtenidos se empleen también para retroalimentar a los estudiantes y para modificar y mejorar la enseñanza.
  4. Que sirvan para verificar el logro de aprendizajes importantes y complejos (competencias) por parte de los estudiantes y que tengan relación con desempeños evaluados mediante la evaluación formativa.

Una de las propuestas sobre evaluación de los aprendizajes que responde –desde la perspectiva de su definición, propósitos, estructuración, instrumentos y medios– a las características antes analizadas es la evaluación del desempeño, conocida también como evaluación de la actuación. Los apartados siguientes de este capítulo se dedicarán a desarrollar sus elementos constitutivos.

EVALUACIÓN DEL DESEMPEÑO

Definición

La definición de la evaluación del desempeño que aquí se ofrece fue elaborada sobre la base de las características establecidas por varios autores:

> "Es un enfoque de evaluación en el que los estudiantes deben ejecutar tareas, mostrar ejecuciones o resolver problemas, en vez de simplemente proveer respuestas marcadas, escritas u orales de una prueba. El desempeño se juzga contra criterios preestablecidos y de naturaleza múltiple. Enfatiza la evaluación de conocimientos y habilidades complejas y de alto nivel de pensamiento, en un contexto del mundo real en que se emplean esos conocimientos y habilidades. Emplea diversos instrumentos y medios que requieren un tiempo sustancial de parte de los estudiantes para completarla" (Wheeler y Haertel, 1993, y Swanson, Norman y Linn, 1995).

Es importante destacar varios elementos de esta definición:

- Los estudiantes realizan tareas, muestran sus logros o resuelven problemas. En otras palabras, se hace énfasis en la producción más que en la reproducción. Son importantes los discursos, la elaboración de objetos y el desempeño que se muestre en productos concretos.
- Se establecen previamente criterios con los que se juzga el trabajo del estudiante. Estos criterios dependerán del juicio de los seres humanos involucrados en el proceso de definirlos.
- Se evalúan conocimientos y habilidades complejas. Ello podría interpretarse como la evaluación de competencias. Esta característica puede implicar el empleo de la indagación disciplinada: aplicación de los conocimientos previos, el entendimiento profundo y la integración para producir, ensamblar, interpretar información, formular nuevas ideas y formular críticas.
- Las tareas, los problemas o los proyectos se contextualizan en la realidad del estudiante y su entorno. Esta característica añade a la evaluación del desempeño la denominación de "auténtica".
- El tiempo que requiere el estudiante para ejecutar las tareas, problemas o proyectos es sustancial. Que puede incluir realizar trabajo fuera de clase y aun por varios días.

La evaluación del desempeño tiene consecuencias para los actores del proceso de enseñanza y del proceso de aprendizaje, por estas razones:

- Es significativa para el estudiante, pues lo convierte en un participante activo de la evaluación. El estudiante debe expresar el conocimiento logrado mediante varias formas de desempeño.
- Al profesor lo obliga a reenfocar su acción didáctica: esta será menos directiva y él se convertirá en un docente más guiador-facilitador del aprendizaje.
- La evaluación se convierte en parte necesaria del proceso de enseñanza y del proceso de aprendizaje, no solo porque la evaluación implica privilegiar la evaluación formativa, sino también porque el estudiante aprende en las actividades de evaluación.
- El currículo se modifica, pues pasa de listados de contenidos y habilidades a descripciones de desempeños, tales como distintas clase de productos, de discursos, de artefactos y de proyectos o tareas que el estudiante debe realizar.
- Los ambientes educativos se transforman, pues se caracterizarán por la colaboración entre estudiantes, por el acceso a recursos de diversa índole, por la flexibilidad en el ritmo del aprendizaje, por la oportunidad de conversar sustantivamente e intercambiar información y por un uso flexible del tiempo.

Medios e instrumentos

Se distinguen varias clases de medios e instrumentos que emplea la evaluación del desempeño:

1. *La observación* del trabajo del estudiante por parte del profesor y por parte del mismo estudiante. La observación será más efectiva si se tiene una guía que dirija la observación y un registro que permita sistematizar las observaciones. También resultan útiles una escala de calificación, una hoja de cotejo o una rúbrica; sin embargo, estos últimos recursos requieren para su aplicación condiciones difíciles de alcanzar en algunos casos.

2. *Muestras de desempeño.* Son documentos o artefactos que indican los logros de los estudiantes. En general, se trata de productos y no de procesos. La valoración de estas muestras requerirá la confección y uso de una rúbrica.

3. *Pruebas.* Estarán constituidas por preguntas de formato abierto, por preguntas analíticas o por problemas. Se trata de pruebas en las que el estudiante podría tener la oportunidad de consultar diversos materiales o que podrían responderse o completarse en su casa.

4. *Tareas o proyectos.* Son trabajos de diversa índole según la disciplina y habilidades de que se trate. Requieren planeamiento, ejecución y la presentación de un informe. Por ejemplo, en Ciencias Naturales podrían estar constituidos por investigaciones sencillas para redescubrir principios y practicar los procesos científicos.

5. *Exposiciones.* Se trata de discursos, obras de arte o artefactos producidos por los estudiantes que se muestran a un público. Pueden ser proyectos tecnológicos o científicos que se exhiben y explican en una feria científica.

6. *Portafolios.* En términos generales, es una carpeta que contiene información acerca de las experiencias y logros del estudiante. Más adelante se realizará una explicación más extensa de este medio.

Las rúbricas

En las explicaciones anteriores se ha establecido que las rúbricas son escalas de calificación que desarrolla el profesor y discute con sus estudiantes. En algunos casos podrían ser elaboradas conjuntamente por el educador y los alumnos. Contienen descripciones detalladas de los niveles de logro que podría alcanzar un estudiante en una tarea, en la elaboración de un producto o de un proyecto. Además presentan una escala, que puede ser cualitativa, cuantitativa o una mezcla de ambas, la cual califica los diferentes niveles de logro. Tal escala está claramente articulada con todos los niveles. Para Mabry (1999), una rúbrica contiene las reglas por las cuales se determina la calidad de las respuestas, los productos o las tareas.

La elaboración de la rúbrica requiere que el profesor tenga claro el marco de referencia conceptual de la tarea que se va a calificar, así como el objetivo o la competencia que se quiere que el estudiante demuestre. Deberá describir los niveles de logro y establecer la escala con que se califica cada uno de esos niveles. Es importante insistir en que la rúbrica debe ser claramente entendida por los estudiantes y que su discusión es necesaria. Más adelante se ofrecerá un ejemplo de rúbrica general para emplearse en Matemáticas.

Ejemplos de evaluaciones de desempeño

A continuación se ofrecerán ejemplos de algunos medios e instrumentos. Estos se tomaron de diversas fuentes, que se señalan en cada uno de ellos.

Matemáticas

TAREA DE TEORÍA DE NÚMERO

Yolanda le contó a su hermano Damián lo que hizo en su clase de Matemáticas.

Yolanda le dijo: "Damián, hoy empleé bloques en mi clase de Matemáticas. Cuando agrupé los bloques en grupos de dos, me sobró un bloque. Cuando agrupé los bloques en grupos de tres, me sobró también un bloque. Y cuando los agrupé en grupos de cuatro, tuve aún un bloque sobrante". Damián le preguntó: "¿Cuántos bloques tenías en total?".

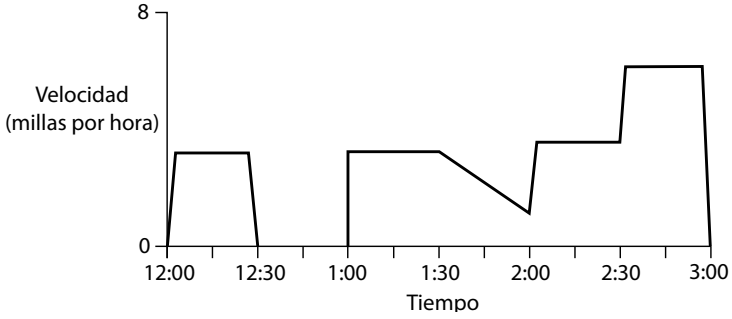
¿Cómo contestó Yolanda la pregunta de su hermano?

Muestre su trabajo

TAREA DE INTERPRETACIÓN DE UN GRÁFICO

Use el gráfico que se muestra más abajo y la siguiente información para escribir una historia o cuento acerca de la caminata de Antonio.

"Al mediodía Antonio empezó a caminar hacia la casa de su abuela, adonde llegó a las 3 pm. El gráfico muestra la velocidad de Antonio, en millas por hora, durante su caminata".



Escriba una historia acerca de la caminata de Antonio. En su historia, describa lo que pudo haber estado haciendo Antonio en los diferentes momentos de su caminata.

Tomado de Lane, S.; Liu, M.; Ankenmann, R. y Stone, C. (1996)

Según los autores, estas tareas están constituidas por preguntas de final abierto y pretenden evaluar un marco de referencia contextual de cuatro componentes: contenido matemático, proceso cognitivo, modo de representación y contexto de la tarea. Este marco de referencia es la base de la rúbrica que se emplea en el proyecto QUASAR (Lane, 1993).

Ciencias

PULSO CARDÍACO

El sistema circulatorio de una persona se ve afectado por muchas razones. El cambio en el ritmo cardíaco puede depender de factores tales como:

- la cantidad de tiempo en que una persona se ejercite (duración),
- la clase de ejercicio que él o ella hace (tipo),
- cuán fuertemente él o ella se ejercite (intensidad) y
- la condición física general de la persona.

Un valor aproximado de la condición física se puede obtener al computar el tiempo que requiere el ritmo cardíaco para regresar después del ejercicio al nivel de descanso. En su proyecto, usted debe hacer lo siguiente:

1. Escribir un problema y una hipótesis acerca de la relación entre el pulso cardíaco de una persona y la duración, tipo o intensidad del ejercicio que él o ella ejecuten. Usted empleará el pulso cardíaco como medida del ritmo cardíaco.
2. Planear, diseñar y ejecutar un estudio para someter a prueba su hipótesis (podría necesitar recolectar datos fuera de clase; sin embargo, todo el trabajo debe completarse dentro de los períodos designados en clase). Recoja datos entre la gente, porque este estudio no debe ser una simple revisión de bibliografía.
3. Tome nota de al menos cuatro resultados. Emplee tablas, cuadros o gráficos para mostrar sus resul tados.
4. Escriba sus conclusiones acerca de la relación, si existe alguna, entre el ritmo cardíaco, la duración, el tipo o la intensidad del ejercicio.
5. Investigue y explique la influencia de la condición física ("fitness") en la(s) relación(es) que usted encontró entre el ritmo cardíaco y el ejercicio. Fundamente sus explicaciones en los datos recolectados y la información que encontró al investigar sobre el tema en la literatura.
6. Escriba un informe final.

Su profesor calificará su proyecto basándose solamente en su informe. Asegúrese de que este incluya:

1. Un problema claramente definido. Escríbalo en forma de pregunta.
2. Una hipótesis que tenga que ver con la relación entre el ritmo cardíaco y la duración, el tipo o la intensidad del ejercicio.
3. Una descripción de su plan y del diseño de su estudio para investigar la hipótesis establecida. Ello responde a la pregunta: "¿cuáles fueron sus procedimientos?". Incluya una descripción de sus variables.
4. Una descripción de la información que recolectó para mostrar la relación entre el ritmo cardíaco y la duración, el tipo o la intensidad del ejercicio. Incluya una discusión de las limitaciones del estudio o de los datos recolectados.

5. Un informe de sus resultados que puede ser fácilmente comprendido por cualquier persona que lo lea. Incluya cuadros, gráficos o tablas si es necesario, así como las relaciones que encontró, si las hay, entre el ritmo cardíaco y la duración, el tipo o la intensidad del ejercicio.

6. Las conclusiones o inferencias que usted obtuvo acerca de la relación entre el ritmo cardíaco y el ejercicio, incluyendo la posible influencia de la condición física.

7. Determine y escriba si su hipótesis era la correcta.

Recuerde que quien lea su informe deberá ser capaz de realizar el mismo estudio usando únicamente la información contenida en tal informe.

Tomado de *Kansas Science Assessment*. Grade 8 (1995)

Esta tarea presenta una gran complejidad para jóvenes de octavo año, pues requiere el uso de conocimientos y habilidades relativas al proceso científico. Los estudiantes deben planificar un estudio completo, establecer problemas e hipótesis, buscar y analizar la información bibliográfica referente al problema, definir procedimientos, recolectar información apropiadamente, analizar la información recolectada, interpretar los resultados a la luz de la hipótesis y de lo que las fuentes bibliográficas establecen acerca de la relación estudiada, y comunicar lo realizado mediante un informe completo. En este informe se fundamentará la evaluación, que se realizará empleando una rúbrica previamente definida.

Lenguaje

TAREA DE ESCRIBIR UNA CARTA

La profesora de cuarto grado planea la enseñanza de la elaboración de una carta.

Primer día. Realiza una discusión con los estudiantes, a los que previamente ha agrupado de cinco en cinco. Les reparte ejemplos de cartas y les pide que en cada grupo se identifiquen las partes que en común tienen las cartas. En la discusión con todos los grupos juntos logra que, con la contribución de todos ellos, identifiquen las partes constitutivas de una carta.

Seguidamente, solicita a cada uno de los alumnos ejecutar en sus casas o en su vecindario una tarea muy concreta. Deberán identificar una situación problemática que tengan en su hogar, en el de algún vecino o en el barrio, que amerite escribir una carta: a algún vecino, al representante vecinal ante el gobierno municipal, a algún departamento municipal, al alcalde municipal, a alguna organización comunal, a alguna organización de servicio (Cruz Roja, bomberos, etc.), a alguna empresa de servicio público (empresas de agua, luz o gas), a alguna autoridad de las iglesias, etc. Tomar nota de la situación, de a qué persona deberán dirigir la carta y anotar el nombre de la persona que informó acerca de la situación problemática.

Segundo día. La profesora solicita a cada estudiante que escriba un primer borrador de la carta que dirigirá a la persona sugerida, describiendo la situación y haciendo una solicitud determinada. Cada estudiante escribe el primer borrador. La profesora les sugiere que hagan una revisión de ese borrador y lo corrijan si lo consideran necesario, y les solicita que, en grupos de cinco, cada uno lea su carta a los demás estudiantes del grupo con el propósito de que reciban observaciones de sus compañeros. Después de este ejercicio, les solicita que corrijan nuevamente con base en las observaciones recibidas de sus compañeros. Recoge esas primeras versiones de las cartas y las fotocopia. Devuelve el original de la carta a cada estudiante y les solicita que la lleven a la persona que les indicó la situación problemática, le pidan

que la lea y que les haga alguna observación relativa a la descripción del problema y a la petición concreta. Les solicita que tomen nota en su cuaderno de las observaciones de dicha persona.

Tercer día. La profesora recolecta los originales y les devuelve la fotocopia de la primera versión. Solicita a los estudiantes que, con base en las observaciones de la persona que les sugirió la situación problemática, escriban una versión final de la carta. La profesora recolecta esta segunda versión.

Una semana después. La profesora devuelve las dos versiones de la carta, con observaciones en cada una, acompañadas de una rúbrica en cada caso. Reúne a los estudiantes en grupos de cinco y les solicita que compartan las observaciones recibidas. Se acerca a cada grupo para escuchar y reafirmar en cada grupo aquellos aspectos que deben ser corregidos para alcanzar mejores resultados. Además solicita a todos los estudiantes que escriban unos párrafos indicando sus sentimientos acerca de la experiencia de aprendizaje vivida y cómo miran ellos el aprendizaje alcanzado, enfatizando los aspectos positivos de lo logrado y cómo se podría mejorar la experiencia de aprendizaje y de evaluación.

Tomado de Esquivel, J. M. (2000)

Nótese que, en el ejemplo de Lenguaje, se trata de una tarea de enseñanza que a su vez se convierte en una tarea evaluada. Este es un ejemplo en que la integración entre enseñanza, aprendizaje y evaluación es total, pues los procesos se funden y se confunden entre sí.

Se muestran a continuación los elementos básicos de la rúbrica utilizada para la evaluación de la tarea de Matemáticas. Además se muestra una rúbrica de Ciencias, que se emplea para la evaluación de una tarea para 5.º grado.

Rúbrica de Matemáticas

NIVEL DE PUNTAJE CUATRO

Conocimiento matemático. Muestra comprensión de los conceptos y principios matemáticos, usa apropiadamente la terminología y notaciones matemáticas, y ejecuta los algoritmos correcta y completamente.

Conocimiento estratégico. Podría emplear información de afuera de naturaleza formal o informal; identifica todos los elementos importantes del problema y muestra comprensión de las relaciones entre ellos; refleja una estrategia apropiada y sistemática para resolver el problema; da evidencia clara del proceso de solución y este resulta completo.

Comunicación. Ofrece una respuesta completa y clara, que incluye una explicación o descripción no ambigua; podría ser un diagrama apropiado y completo. Comunica eficazmente a la audiencia identificada; presenta argumentos fuertes y con soporte que son lógicos y completos; podría incluir ejemplos que ilustren la respuesta y ejemplos contrarios.

NIVEL DE PUNTAJE CERO

Conocimiento matemático. No muestra comprensión de los conceptos y principios del problema matemático.

NIVEL DE PUNTAJE CERO

Conocimiento matemático. No muestra comprensión de los conceptos y principios del problema matemático.

Conocimiento estratégico. Podría intentar emplear información de afuera irrelevante; falla en indicar cuáles elementos del problema son apropiados; copia parte del problema, pero no intenta ninguna solución.

Comunicación. Se comunica ineficientemente; las palabras que emplea no reflejan el problema; podría incluir dibujos que no representan la situación que establece el problema.

Existen descripciones para las tres categorías en los niveles de puntaje tres, dos y uno, que no se transcriben por problemas de espacio. Pueden consultarse en Lane, S. (1993) o en Lane, S.; Lui, M.; Ankenmann, R. y Stone, C. (1995).

Rúbrica de Ciencias

COMPONENTES DEL PROCESO

Reconoce y define el problema.

Diseña la estrategia de resolución del problema.

Implementa la estrategia de resolución del problema.

Interpreta y comunica resultados y conclusiones.

Los cuatro componentes del proceso se califican en una escala de cinco valores; por ejemplo, para el componente "Reconoce y define el problema".

Puntaje uno: La observación y la descripción de objetos y sucesos es limitada. Escribe una pregunta que no es aceptable o no escribe la pregunta.

Puntaje dos: Entre 1 y 3.

Puntaje tres: Observa y describe algunas características singulares de los objetos o los sucesos. Escribe preguntas aceptables, pero no claramente definidas.

Puntaje cuatro: Entre 3 y 5.

Puntaje cinco: Observa y escribe características singulares de los objetos o fenómenos. Escribe una pregunta aceptable para ser objeto de investigación.

Tomado de *Kansas Science Assessment*, *Fifth Grade, Teacher Manual*, 1995

Portafolios

De las diversas opciones de instrumentos y medios que presenta la evaluación del desempeño, se realiza una explicación detallada del portafolio por ser uno de los medios más empleados. Muchas veces se hace sin respetar los elementos básicos y los propósitos que tiene este medio.

Definición

La siguiente definición de portafolio fue elaborada sobre la base de los elementos ofrecidos por los autores que se señalan más adelante.

> "Es una colección comprensiva del trabajo del estudiante que exhibe sus esfuerzos, progresos y rendimientos en una o más áreas. La colección debe incluir la participación del estudiante en seleccionar

> los contenidos, los criterios de selección empleados, los criterios para juzgar el mérito de su trabajo y la demostración de autorreflexión sobre el trabajo. Las descripciones de sus logros pueden incluir muestras de grabaciones de audio, vídeo y fotografías que registren desempeños y proyectos, y testimonios de autoridades sobre la calidad del trabajo del estudiante. En resumen, contiene una variedad de información que documenta las experiencias y logros del estudiante" (Paulson, F.; Paulson, P. y Mayer, C., 1991, y Archabald, D. y Newman, F., 1991).

Agregan Paulson, Paulson y Mayer (1991): "Un portafolio es un portafolio cuando provee una visión contextualizada, comprensiva y compleja del desempeño del estudiante. Es un portafolio cuando el estudiante es un participante en vez de ser objeto de evaluación. Pero, sobre todo, un portafolio es un portafolio cuando provee un foro que estimule al estudiante a desarrollar las habilidades necesarias para llegar a ser un aprendiz independiente" (p. 15).

De la definición anterior y de lo expresado por diversos autores se pueden destacar algunos aspectos importantes:

- Al desarrollar un portafolio, se ofrece al estudiante una oportunidad para aprender acerca del aprendizaje.
- El portafolio es un producto hecho *por* el estudiante y no *para* el estudiante.
- El portafolio difiere de una carpeta acumulativa sobre un estudiante.
- Podría tener un carácter formativo durante el año y un carácter sumativo al final del año.
- Podría cumplir propósitos múltiples, pero estos no deben estar en conflicto.
- Debe dar a entender explícita o implícitamente las actividades del estudiante, pero también debe mostrar todos sus logros.
- Debe contener información que ilustre el crecimiento del estudiante.
- Los estudiantes muchas veces requieren modelos y ejemplos de portafolios para producir sus propios portafolios eficientes.

Portafolio sobre producción escrita

Un portafolio que es, a su vez, una selección realizada por el estudiante del contenido del portafolio, que el estudiante ha elaborado durante todo el año, acerca de su capacidad de producción escrita. Eventualmente, el profesor selecciona también alguna producción.

Al estudiante se le pidió que seleccionara cuatro piezas, calificadas así:

– Una pieza importante para él.

– Una pieza que le fue satisfactoria.

– Una pieza no satisfactoria para él.

– Una pieza de selección libre.

– Ocasionalmente, el profesor seleccionaba otra pieza .

Además debe presentar una tabla de contenidos. Un inventario de la producción realizada en el año y una descripción de su experiencia como escritor. Una reflexión final sobre su producción que ofrezca una evaluación personal y la explicación de los cambios que han experimentado durante el año su producción y su capacidad como escritor.

La rúbrica que se empleó tenía tres dimensiones:

– Logros en la producción escrita.

– Uso de procesos y estrategias para escribir.

– Crecimiento, desarrollo y compromiso como escritor.

Cada una de estas dimensiones está acompañada por una descripción de elementos constituyentes que indicaban criterios importantes para calificar la dimensión (no se detallan por falta de espacio).

Cada una de estas dimensiones se calificaba en una escala de siete puntos: desde el puntaje seis, que indicaba excelente desempeño, hasta el puntaje cero, que indicaba desempeño totalmente inadecuado. Cada dimensión se hace mediante una calificación holística.

LeMahieu, P.; Gitomer, D. y Eresh, J. (1995)

CONCLUSIÓN

Los profesores habrán encontrado en este capítulo la justificación sustentada de la necesidad de un paradigma de evaluación que sea capaz de integrar el trinomio enseñanza-aprendizaje y evaluación. También la fundamentación de una evaluación que tenga una función pedagógico-didáctica, que sea consonante con la teoría de aprendizaje socio-contructivista. La evaluación que se ha propuesto para que se emplee en las aulas se caracteriza por: basarse en objetivos y criterios claros, ser participativa, emplear una interpretación criterial, ser formativa y formadora, porque sus medios e instrumentos obligan a los estudiantes a realizar procesos mentales complejos y retadores, retroalimentar e indicar cómo superar las deficiencias, requerir el diseño de actividades de evaluación muy semejantes a las de enseñanza, ser continua y porque sus resultados permiten mejorar la enseñanza.

La visión general sobre la evaluación del desempeño, la explicación de sus medios e instrumentos y los ejemplos que se han ofrecido, se espera que se constituyan en un reto para que los profesores investiguen las fuentes bibliográficas y encuentren ideas que les permitan mejorar la evaluación de los aprendizajes. Finalmente, es importante señalar que la evaluación del desempeño ofrece la fundamentación, los medios e instrumentos para ser una de las mejores opciones para la evaluación de las competencias.

El sentido de la evaluación

Difusión y uso de resultados de evaluaciones educativas a gran escala en América Latina

Patricia McLauchlan de Arregui

INTRODUCCIÓN

Así como durante la etapa de expansión del acceso a la educación en Latinoamérica se dio prioridad a la recogida y utilización de indicadores de cobertura y niveles de escolaridad, cuando lograr la efectividad de las escuelas y el aprendizaje real se convirtió en el principal desafío a finales de los ochenta, y se tuvo que diseñar y justificar nuevas estrategias para la asignación de recursos, la medición de la calidad de la educación y los aprendizajes se hizo impostergable.

Así, a mediados y finales de los años noventa la mayoría de países de América Latina empezó a dar sus primeros pasos hacia el desarrollo de capacidades técnicas básicas para diseñar e implementar evaluaciones a gran escala y para emprender análisis estadísticos de los datos recogidos. Una visión general del desarrollo de los sistemas de evaluación de los logros de aprendizaje puede verse en Ferrer (2006).

Hoy, tras varios años de experiencias, en buen número de países se están dando discusiones y acciones orientadas al rediseño de los sistemas de evaluación. Una parte del debate se refiere a necesarias mejoras técnicas en el desarrollo de instrumentos y procedimientos para la recolección y el análisis de datos, así como al alineamiento de los mismos con los objetivos curriculares vigentes. Otra parte tiene que ver con la constatación de que los esfuerzos para comunicar y difundir los resultados de las pruebas no han alcanzado a muchos agentes clave, incluyendo profesores y padres (Ravela, 2002). Tanto o mayor desasosiego genera el hecho de que, aparentemente, la abundante información que se está produciendo está siendo usada de modo insuficiente para sustentar y apoyar procesos de mejora, tanto en las políticas como en la enseñanza, o como instrumento de responsabilización o *accountability*. Las evaluaciones parecerían, en opinión de algunos, no estar teniendo mucho impacto *más allá de informar, sensibilizar y crear conciencia* (Moncada *et al.*, 2005). Por supuesto, no faltan quienes están más preocupados aún por los usos no válidos o hasta perversos de las evaluaciones para el logro de objetivos para los cuales no fueron diseñadas.

La primera sección de este capítulo revisa de manera resumida la experiencia emergente de los países de América Latina en la difusión de resultados de evaluaciones, mientras que la segunda pasa revista al uso de los resultados de evaluaciones nacionales, sub-nacionales, regionales o internacionales. Por último, se presentan algunas generalizaciones y advertencias sobre los desafíos que supone una utilización más productiva de los resultados.

LA DIFUSIÓN DE RESULTADOS[1]

La mayoría de los sistemas nacionales de evaluación desarrollados en América Latina durante los noventa enfatizaron sus finalidades de carácter formativo y sin consecuencias directas sobre las escuelas. Sus roles explícitos fueron generalmente constituirse en proveedores de información para la toma de decisiones por parte de las autoridades y actores del sector educativo, así como para el público sobre el estado de la educación nacional. Su cumplimiento requería, por lo tanto, *difundir* y *utilizar* la información recogida para rediseñar o ajustar políticas, planes, programas y prácticas pedagógicas y de gestión escolar. La concreción de esas intenciones, sin embargo, no ha estado exenta de una serie de limitaciones.

En primer lugar, falta hasta ahora en muchos Ministerios una *visión estratégica sobre el rol de la evaluación*. No se tiene clara la manera en que se espera que esta impacte sobre el sistema educativo. En casos extremos, parecería creerse que el mero hecho de evaluar producirá mejoras o que todos los actores comprenderán y emplearán los resultados que se divulguen. Hay unidades y actores de medición de logros educativos que consideran que su misión consiste únicamente en producir datos e información cuyo análisis no les corresponde y debe ser realizado por otras agencias o actores, pasando por alto que en realidad aún no existen muchas audiencias en condiciones de comprender y utilizar la información, por lo cual es necesario invertir tiempo en formarlas y en "modelar" análisis sólidos y válidos de los datos.

Falta aún en algunos países una definición clara de cuáles son las audiencias a las que se debe y quiere llegar, cuáles son los contenidos y mensajes que se quiere hacer llegar a cada una y cuáles son los vehículos o tipos de productos de difusión apropiados en cada caso. Incluso hoy día no es del todo extraño que se produzca un único informe de carácter predominantemente técnico, pensado desde la perspectiva de los especialistas de la unidad de evaluación y no de las demandas y preocupaciones de las audiencias, ni siquiera de las audiencias internas del sector educativo, obviando también a las familias, la opinión pública, el sistema político, el mundo académico, las organizaciones sociales y hasta a la prensa. En los casos en que sí se procura llegar con informes a estos otros actores, generalmente no se ha evaluado el grado de comprensión y uso que ellos hacen de la información recibida.

Otro rubro de problemas que ha afectado a la difusión de los resultados ha resultado de inconsistencias técnicas y desarticulaciones internas que caracterizaron las actividades iniciales de las agencias evaluadoras.

Un ejemplo de ello es la confusión generada por algunos Ministerios en torno a pruebas que son de tipo descriptivo y diagnóstico, pero que son interpretadas por la prensa y el público en términos de aprobación y reprobación de alumnos. La demanda pública parece ir en un sentido –y quizá deba ser educada para que asuma otra orientación–, pero las pruebas fueron diseñadas desde otra perspectiva, lo cual no permite satisfacer esa demanda.

Otro problema crucial que ha dificultado la difusión y aprovechamiento es la falta de articulación y comunicación entre las unidades de evaluación y de currículo al interior de los Ministerios. Suele ocurrir que las pruebas no estén adecuadamente alineadas con los objetivos establecidos

[1] Esta sección está basada fundamentalmente en el documento *¿Cómo aparecen los resultados de las evaluaciones educativas en la prensa?* del Grupo de Trabajo sobre Estándares y Evaluación del PREAL.

en el currículo o que estos últimos estén formulados en términos tan vagos y generales que es imposible realmente evaluar su logro. En estos casos, las evaluaciones no pueden responder a la pregunta principal de si los alumnos están aprendiendo lo que se espera que aprendan y qué tan bien o mal están al respecto. En su capítulo, Elena Martín analiza con más detalle las relaciones entre currículo y evaluación de rendimiento.

No menos grave ha sido el uso político de los resultados por los propios Ministerios. En parte debido a una suerte de visión ingenua que han tenido algunas altas direcciones de esas instituciones respecto a que los resultados pueden mejorar rápidamente, ellas no han estado muchas veces preparadas para manejar con solvencia los malos resultados ante la prensa. Al respecto se han observado diversas reacciones: a) las autoridades simplemente deciden no publicar los resultados o, lo que es más común en años recientes, los publican casi "subrepticiamente", sin una estrategia integral de difusión; b) al no dominar el tema, transmiten mensajes que incluyen errores de interpretación importantes; c) utilizan los malos resultados para desprestigiar a la administración anterior (normalmente al inicio de un mandato) o para presionar sobre el cuerpo docente, aunque luego se esperará que los mismos mejoren rápidamente antes del término de su gestión. Como dice Pedro Ravela, "normalmente los funcionarios políticos esperan buenos resultados que les sirvan como propaganda o aval a su gestión, y los malos resultados, que suelen ser los más comunes, se tornan una 'papa caliente' que no saben cómo manejar" (2003).

Un tema vinculado a los anteriores es el dilema de la independencia política de las unidades de evaluación. Estas producen y divulgan la información sobre logros educativos, pero en muchos casos son a la vez parte interesada, lo cual coloca inevitablemente a la información bajo sospecha para los periodistas y otros actores. Esta falta de confianza se da incluso cuando la unidad de evaluación está fuera de la estructura ministerial, ya que igualmente dependen de los recursos del Estado para realizar las evaluaciones y de decisiones políticas para difundirlos.

El problema no radica tanto en el marco institucional elegido como en la cultura política y en la cultura de la evaluación. Es necesario –y en muchos países se ha logrado avances en ese sentido– que los Ministerios de Educación conciban la labor de evaluación como una de información al público y otorguen a las unidades de evaluación amplia independencia en la difusión de los resultados. Igualmente se requiere que la sociedad se acostumbre a que existe una unidad de evaluación que periódicamente informa sobre el estado del sistema educativo y de los aprendizajes de manera clara, técnicamente apropiada y sin interferencias de tipo político. Ello no impide, por supuesto, que los Ministerios deban tener y ofrecer su propia visión e interpretación de los resultados y debatan con otros sectores de opinión al respecto.

Un último grupo de problemas tiene que ver con las débiles relaciones con los medios de comunicación que mantienen algunas unidades evaluadoras. Estas parecen a veces desconocer la lógica y el *modus operandi* de los medios y no haberse trazado estrategias ni dedicar recursos para vincularse de mejor manera con ellos. Por su parte, el personal de prensa de los Ministerios está generalmente dedicado principal o casi exclusivamente a organizar y difundir las apariciones públicas de las autoridades y a cuidar su imagen, antes que a facilitar la comunicación con los medios sobre los temas educativos.

Frente a esta situación, el Grupo de Trabajo sobre Estándares y Evaluación del PREAL ha recomendado que las agencias de evaluación propongan, y los Ministerios definan, *qué preguntas pretenden responder las evaluaciones*, qué tipo de información se va a producir, quiénes y cómo se

espera que utilicen la información. Luego, que *definan una estrategia de difusión de resultados,* identificando diferentes audiencias y sus respectivas preguntas, así como el tipo de productos de comunicación apropiados para cada una. Se recomienda informar siempre de modo transparente, sin interferencias políticas, y entregar la información a actores cualificados con diferentes posturas, a los cuales pueda recurrir la prensa para obtener interpretaciones y visiones diferentes a partir de los resultados, evitando la tentación de usar los resultados con fines propagandísticos o de ocultarlos cuando no son satisfactorios.

Es recomendable que la misma unidad de evaluación explique los resultados, incentivando a las autoridades del sector a que reflexionen e informen sobre las implicancias que tienen para las políticas educativas. Ello requeriría que se preparen mensajes claros, balanceando buenas y malas noticias, evitando tanto el catastrofismo como el triunfalismo, presentando los resultados como desafíos al sistema, *como un diagnóstico y no como autopsia* (Ravela, 2003), y enfatizando las líneas de acción con que se plantea enfrentar los problemas que se ponen de manifiesto.

Para fomentar la apropiación de los resultados de la evaluación, es necesario que actores de otras oficinas del Ministerio se sientan también propietarios del proceso, participando en el diseño y validación de las pruebas, así como en la interpretación de resultados. Contribuiría también a la aceptación transmitir que el sistema de medición seguramente tiene debilidades, pero que está abierto a mejoras, para lo cual se quiere recoger la opinión cualificada de actores diversos, incluyendo a la prensa, a la cual se quiere convertir en aliado crítico.

Es conveniente que los mensajes sean pocos, claros, reiterativos y basados en la evidencia resultante de las pruebas y del análisis de los factores vinculados. Debe ejemplificarse el tipo de conclusiones que no pueden ser derivadas de los datos y las principales limitaciones de la evaluación, a la vez que se preparan explicaciones pertinentes y al alcance de diferentes audiencias acerca de aspectos técnicos centrales, como la equivalencia (o no) de las pruebas con mediciones anteriores; el significado de las métricas empleadas; la posibilidad (o no) de interpretarlas en términos de *aprobación*; el modo en que se han establecido *líneas de corte* (si se ha hecho), o el significado de los *niveles* (si se han establecido).

El tenor de los mensajes debería ser que la educación es una política de Estado y un compromiso de toda la sociedad, evitando tanto la búsqueda de culpables o responsables únicos (los docentes, las escuelas, los supervisores) como el uso de las acciones educativas con fines de propaganda política, informando con modestia y cautela sobre los avances y los desafíos pendientes.

En términos de *las relaciones con la prensa,* es necesario asegurar la disponibilidad de recursos humanos idóneos y suficientes para la labor de comunicación con ella. Deberán organizarse actividades de capacitación para periodistas con relación a los temas clave de las evaluaciones, idealmente en momentos separados de la presentación de resultados. Muchas de las explicaciones y conceptos técnicos, así como de las advertencias sobre lo que no puede concluirse válidamente de los datos o el tipo de interpretaciones que no son apropiadas, pueden resultar "sospechosos" para el periodista cuando se hacen en el marco de la difusión de resultados. Convendría a las unidades de evaluación cultivar vínculos con periodistas reconocidos o especialmente interesados en temas educativos, y mantenerlos permanentemente informados, no con el objetivo de contar con "incondicionales acríticos", sino más bien con "incondicionales acerca de la importancia de la difusión de resultados y del debate público sobre la educación", capacitándolos en modos de hacer "noticia educativa" de manera constructiva.

LOS USOS DE LOS RESULTADOS

Los resultados de las evaluaciones se han utilizado para varios propósitos, como sigue:

Elevar la conciencia ciudadana sobre los retos de la educación

Si bien las evaluaciones del rendimiento se iniciaron generalmente a partir de la necesidad de informar a los tomadores de decisiones, el contexto de resurgimiento democrático de la época y la idea de *rendición de cuentas* que empezaba a instalarse en el discurso político también hacía necesario satisfacer los derechos de los ciudadanos y de algunas audiencias clave de conocer la situación y las tendencias de los resultados educativos. Los resultados han provisto de pruebas duras sobre los logros insuficientes y desiguales, han estimulado una preocupación pública sobre el problema y reorientado el foco de la demanda social hacia una distribución más equitativa de oportunidades de aprendizaje.

Aun así, en la mayoría de los casos la prensa recogió de los reportes oficiales solo los resultados más llamativos sobre el bajo nivel de logro general, y en el caso de las pruebas internacionales, sobre los bajos puntajes obtenidos por los participantes latinoamericanos. A menudo se puso más énfasis en las grandes diferencias entre las escuelas privadas y las públicas, y mucho menos en aquellas entre las poblaciones urbana y rural o entre niñas y niños. Rara vez se mencionaron las diferentes características socioeconómicas y culturales de las poblaciones atendidas por escuelas privadas y públicas. Por otro lado, al menos al inicio, los funcionarios públicos rara vez estaban disponibles para analizar los resultados y promover un diálogo capaz de orientar a la opinión pública a partir de las evaluaciones.

Entre las audiencias interesadas son quizá los padres de familia el grupo menos atendido, en lo que respecta a proveerles con información útil para monitorear el desempeño de las escuelas a las que asisten sus hijos o del sistema educacional en su conjunto. Los reportes producidos están más allá de las posibilidades de comprensión de la mayoría de los padres. El manejo de la información por la prensa no ha ayudado a que la información se interprete de forma precisa, y las capacidades de traducción de los sofisticados procedimientos técnicos de evaluación y análisis estadístico son aún limitadas.

Mejorar el diseño de las políticas educativas y la asignación de recursos

Gracias a las evaluaciones nacionales e internacionales, se cuenta con datos más sólidos sobre el bajo rendimiento de los estudiantes, así como sobre las grandes diferencias en los resultados de varios grupos poblacionales y tipos de escuelas, lo que ha hecho posible el establecimiento y la generación de consensos sobre nuevas prioridades generales de política.

La necesidad de que los Ministerios de Educación, los administradores y las escuelas se centren en el aprendizaje de los estudiantes es ahora un mandato público generalizado en la región. Y, si bien la focalización y la discriminación positiva o acción afirmativa están hasta ahora lejos de ser el eje principal de las políticas públicas, existe una creciente conciencia de la necesidad de distribuir más equitativamente las oportunidades de aprendizaje.

Muchas políticas inspiradas por los resultados de las evaluaciones no han logrado trascender el papel donde han sido formuladas; en cambio sí se ha pasado a la acción con reformas curriculares y desarrollo de estándares, producción y distribución masiva de libros de texto y cuadernos de

trabajo, el fomento de la inclusión de estudiantes pobres en escuelas particulares subvencionadas, programas de capacitación o desarrollo profesional docente focalizados en el fortalecimiento de capacidades de comprensión lectora y de razonamiento matemático, o subvenciones diferenciadas que favorezcan a las escuelas y estudiantes más pobres.

Diseñar programas de apoyo y seleccionar escuelas o estudiantes para participar

Las evaluaciones han permitido identificar áreas más específicas donde hay problemas, analizar sus posibles causas y diseñar programas para superarlos, así como identificar y seleccionar poblaciones objetivo o instituciones educativas que puedan ser beneficiarias.

Un ejemplo de estos programas se dio en Chile, cuando los resultados de las pruebas SIMCE mostraron grandes diferencias de rendimiento entre escuelas privadas, subvencionadas y municipales y se dio inicio como una estrategia de emergencia al P-900, el Programa para 900 escuelas *Ayudando a los pobres*. El 10% de escuelas con el rendimiento más bajo en cada una de las regiones del país, con grandes proporciones de estudiantes pobres urbanos y rurales de nivel primario, fue invitado a participar en el programa, inicialmente concentrado en remodelar aulas y abastecer de equipo básico, libros de texto y bibliotecas a las escuelas, pero luego incluyendo talleres para profesores que buscaban promover la autonomía profesional y la responsabilidad, y espacios de tutoría en grupos pequeños para estudiantes con bajo rendimiento. El P-900 se convirtió más adelante en un programa permanente, centrado en proveer servicios de capacitación de profesores *in situ* y desarrollo profesional o capacitación en servicio de directores a lo largo de tres años. Empezó a incluir también escuelas de nivel secundario y abrió un subprograma especial para escuelas rurales.

El programa se fue modificando conforme la investigación y la práctica iban mostrando más elementos que necesitaban ser atendidos para promover cambios, manteniéndose la prioridad del desarrollo de capacidades fundamentales en lectura, escritura y matemáticas. Las escuelas P-900 –que llegaron a sobrepasar las 1.500– fueron el foco de programas específicos de mejora y eran supervisadas entre 8 y 16 veces al año por funcionarios provinciales y centrales.

Similares en enfoque y objetivos han sido la iniciativa de *Nivelación para la Excelencia* de Bogotá, las escuelas de jornada completa de Uruguay establecidas preferentemente en áreas urbanas empobrecidas y surgidas de la constatación que ellas habían obtenido mejores logros en la primera evaluación nacional, y la jornada completa introducida también en Chile.

Las políticas de focalización se han convertido en una característica frecuente en muchos países de América Latina. Esto requiere indicadores técnicos sólidos y socialmente aceptables con los cuales identificar las poblaciones objetivo de programas especiales de apoyo, y las evaluaciones se han constituido en un insumo importante en este sentido.

Realizar investigación y evaluar el impacto de programas

Si consideramos que el aprendizaje es la misión central de la educación, es posible afirmar que la variable dependiente más importante para la investigación educativa orientada a la formulación de políticas no estaba disponible en gran parte de Latinoamérica antes de que los sistemas de evaluación comenzaran a generar resultados. Y dado que la investigación debería ser un insumo importante para la toma de decisiones, la región mostraba un serio retraso.

Hasta 1999 en toda América Latina no se podía encontrar más de una veintena de estudios empíricos sobre posibles asociaciones entre las diferencias en rendimiento, diferencias en las características intraescolares o de aula y diferencias en factores externos tales como las características educativas y la condición socioeconómica de las familias. Apenas seis años más tarde una consulta bibliográfica rápida reveló la existencia de 150 estudios en 16 países, que exploran los efectos específicos o conjuntos sobre el rendimiento de factores tales como: a) la condición socioeconómica de los estudiantes, el trabajo durante el año escolar, la raza y la condición nutricional; b) el salario, calificaciones y prácticas de enseñanza en aula de los docentes; c) características de la escuela tales como la cultura organizacional, la autonomía, la composición socioeconómica de los distintos grupos, las prácticas de evaluación en el aula, la cobertura del plan de estudios, las políticas sobre la realización de tareas en casa, el liderazgo de los directores, la infraestructura, la disponibilidad de libros de textos, y d) el financiamiento o administración pública o privada de la escuela, la presencia de programas de bonos escolares (*vouchers*), de programas de alimentación escolar, las prácticas de retención/repetición, el número de alumnos por aula y la duración de la jornada escolar.

Los resultados de las pruebas se están también utilizando cada vez más para estimar el impacto de políticas y programas educativos y sociales. Los Ministerios incluyen muestras mayores de algunas poblaciones escolares, de manera de poder hacer algunas generalizaciones válidas sobre el impacto de intervenciones especiales, tales como escuelas bilingües indígenas (en Perú y Guatemala), programas de expansión escolar basados en la comunidad (*Nueva Escuela Unitaria* de Guatemala, escuelas PRONADE), perspectivas nuevas o experimentales para planes de estudios (*Escuelas de Excelencia* y el SIMAC en Guatemala) y mecanismos de gestión y financiamiento (escuelas privadas subvencionadas en Chile y Nicaragua). Con frecuencia también se ha permitido e incluso se ha estimulado a los investigadores a usar los resultados de las pruebas nacionales o re-aplicaciones o adaptaciones de los instrumentos de las pruebas para evaluar el impacto de programas específicos.

Revisar expectativas curriculares o establecer estándares de aprendizaje

Hoy se entiende que elevar y aclarar las expectativas de aprendizaje constituye un elemento crucial para su mejora. Si ellas no están bien definidas, no pueden ser utilizadas para alinear de forma efectiva los libros de texto, los planes de estudio y la capacitación docente.

Cuando los sistemas de evaluación en América Latina empezaron a desarrollar pruebas referidas a criterios, experimentaron grandes dificultades para definir los tipos de tareas que se debería esperar que los estudiantes puedan realizar, si han logrado dominar un área particular de los contenidos. También confrontaron –y muchos siguen confrontando– grandes desafíos para definir los niveles de desempeño en cada criterio seleccionado. Allí donde se ató alguna consecuencia a los resultados de esas evaluaciones, o donde se han realizado esfuerzos importantes por devolver resultados a las escuelas, los contenidos de las pruebas han empezado a cumplir la función de *aclaradoras* del currículo, si bien no todos los curriculistas aprobarían el que ello sea así. La amplitud de las consultas a docentes, expertos y productores de textos que pueda realizar la agencia evaluadora también ha contribuido en algunos casos a generar un proceso de redefinición de metas curriculares nacionales.

Mayor ha sido la influencia de las pruebas internacionales. Los resultados de los países que han participado en TIMSS, PISA y otras evaluaciones internacionales no siempre fueron congruentes

con la distribución de estudiantes entre los distintos niveles de rendimiento fijados en sus propias pruebas nacionales o con los índices de aprobación[2]. Esto los impulsó a que dieran una nueva mirada a sus currículos buscando posibles explicaciones y, en algunos casos, a profundizar sus esfuerzos de aclarar metas de aprendizaje estableciendo estándares.

Los primeros esfuerzos fueron desplegados por los países centroamericanos en 1999, bajo la conducción de la Secretaría General del organismo de Coordinación Educacional y Cultural de la región. Desde entonces, Honduras y Guatemala han avanzado sendas propuestas de estándares para todos los grados de la educación básica, para lo cual han tomado en consideración los resultados logrados en evaluaciones nacionales anteriores, y con los cuales empiezan ya a alinear un conjunto de instrumentos de evaluación de aprendizajes. En Nicaragua, el conjunto de estándares de la SGCECC fue utilizado para revisar el currículo y sirvió de insumo para la construcción de su primera evaluación nacional, aplicada en el 2003.

Colombia desarrolló y estableció estándares en el 2003, tomando como uno de sus referentes la información disponible sobre las capacidades actuales de los estudiantes según lo evaluado por el Instituto Colombiano de Fomento de la Educación Superior (ICFES). Los estándares se centran en competencias para resolver problemas de tipo real en lenguaje, matemáticas, ciencias naturales y sociales, y más recientemente en ciudadanía. Se espera que sirvan como guías para el desarrollo de programas de mejora y que se conviertan en el conjunto básico de expectativas de aprendizaje que el ICFES utilizará para las evaluaciones en corto plazo.

En Chile, el análisis del SIMCE y de los resultados de las pruebas internacionales ha mostrado poco incremento significativo en los puntajes promedio, a pesar de los grandes aumentos en el gasto público en educación, de las reformas curriculares y de la inversión en varios proyectos a gran escala. Se introdujeron sucesivamente varios nuevos instrumentos y aclaraciones curriculares, antes de optar por la construcción de estándares de contenido y desempeño, conocidos ya como Mapas de Progreso y Niveles de Logro, respectivamente. Los Mapas y Niveles se nutrieron, entre otros elementos, de los resultados obtenidos por los alumnos chilenos en sus pruebas nacionales e internacionales. Se convertirán gradualmente en referentes básicos para la elaboración de los instrumentos de evaluación y se usan ya en el reporte de resultados del SIMCE. Se pretende que sirvan también para guiar la reflexión, la planificación y la práctica pedagógica de los docentes, en subáreas específicas de matemáticas, lenguaje, ciencias sociales, ciencias naturales e inglés como segunda lengua y, más adelante, en educación física, educación preescolar, educación técnica y arte.

Promover la responsabilización, rendición de cuentas y libre elección de escuelas

En general, las pruebas estudiantiles con altas consecuencias (aprobación, graduación de alumnos, ascenso o salida de docentes, cambio de autoridades o cierre de escuelas) no son comunes en América Latina, a excepción de pruebas de tránsito a niveles superiores del sistema educativo como en Costa Rica, República Dominicana y, más recientemente, en El Salvador. Sin embargo, también tienen consecuencias las pruebas nacionales cuyos resultados sirven para seleccionar a escuelas para recibir premios o sanciones, ya sean simbólicos o efectivos, tales como la publi-

[2] Enrique Roca, en su capítulo en este libro, revisa con detalle estas evaluaciones.

cación de *rankings*, premios a la excelencia de instituciones, incentivos para equipos docentes o ascensos individuales en la carrera magisterial.

La práctica más simple y extendida a este respecto es *la identificación de las escuelas más destacadas en rankings nacionales* que se hacen públicos ya sea intencionalmente o por iniciativa de los medios, como en Chile, El Salvador o Guatemala. En Colombia se otorga el Galardón a la Excelencia a las escuelas cuyos resultados en las pruebas presentan mayores mejoras, mientras que las escuelas públicas más destacadas son invitadas a participar de una Red de Acción para la Excelencia a través de la cual comparten información sobre sus prácticas exitosas con otras escuelas.

Se argumenta muchas veces que esta práctica suele convertirse en un incentivo a los *ganadores* para lograr progresos adicionales antes que en un estímulo a los que se encuentran en los niveles más bajos de logro. Los educadores en América Latina, como los de muchos otros países del mundo, suelen estar fundamentalmente opuestos a la publicación de *rankings* y consideran que los resultados de las pruebas solo deberían compartirse al interior de la comunidad escolar. Muchos expertos en evaluación lamentan además la manera poco refinada en que se construyen la mayoría de esos *rankings*, que proyectan como diferencias reales en calidad diferencias entre puntajes que pueden atribuirse a errores de medición.

Otro tipo de uso que se da a los resultados de las evaluaciones son *los incentivos al desempeño docente y escolar*. Dadas las dificultades para identificar exactamente qué insumos educativos "garantizan" mejores resultados de rendimiento en distintas condiciones, los diseñadores de políticas analizan varias combinaciones de incentivos y mecanismos de responsabilización con los cuales movilizar actores e instituciones clave en su propia búsqueda de las "mejores" direcciones a seguir. La investigación sobre el tema, si bien muestra que las compensaciones económicas no son el más eficaz incentivo al desempeño docente, sugiere también que estas pueden servir para tal propósito cuando se las incluye dentro de una estrategia mayor para asegurar que los profesores concentren sus esfuerzos en el rendimiento estudiantil y, particularmente, cuando requieren de un esfuerzo colectivo para lograr los premios (Odden y Kelly, 1997).

Aun así, evaluaciones de programas tales como el Sistema Nacional de Evaluación del Desempeño de Docentes (SNED) en escuelas con financiamiento público de Chile, o Carrera Magisterial en México, que han utilizado con diferente peso pero de manera directa los resultados de las evaluaciones estudiantiles como criterio de adjudicación de un incentivo, no parecen haber tenido los efectos buscados (Mizala y Romaguera, 2004, y Santibáñez *et al.*, 2007) y sí, en cambio, algunas consecuencias indeseables aunque no intencionales. Las limitaciones de estos y otros esquemas de incentivos similares basados en resultados (como indicadores de desempeño) se discuten en el capítulo de Pedro Ravela en este volumen.

También se han usado los resultados de las evaluaciones para facilitar que *los padres puedan elegir libremente a qué escuela enviar a sus hijos*, lo que de forma indirecta se supone que sirve para presionar a las escuelas a mejorar la calidad de sus servicios y sus resultados. Esto es así particularmente en aquellos lugares donde el financiamiento público se asigna en función del número de niños atendidos, como en el caso de los sistemas de subvenciones escolares en Chile. Los directores de cada escuela reciben un informe de su desempeño en cada una de las áreas curriculares que han sido evaluadas, comparándolo con el de escuelas similares y promedios nacionales, así como información sobre las tendencias observadas en el tiempo a este respecto. Además se pone a disposición del público general los resultados obtenidos en las pruebas por cada una de

las escuelas de cada región del país, junto con los promedios para cada área, información sobre los cambios observados desde las rondas anteriores y comparaciones con escuelas con características socioeconómicas similares. Se elaboran informes para la prensa y para los padres, buscando revertir la tendencia de la prensa a construir clasificaciones simplistas sobre cuáles son las *mejores o peores* escuelas o a publicar titulares escandalosos sobre la ausencia de mejoras en el rendimiento.

Las investigaciones acerca de si el acceso a la información sobre el rendimiento sirve realmente para ejercer una presión efectiva sobre las escuelas para mejorar su oferta no son concluyentes (Elacqua y Fábrega, 2004). No se sabe realmente si el esquema competitivo ha servido o no para estimular mejoras en el sistema ni en las escuelas municipales subvencionadas que atienden a los pobres, mientras que parecería que las escuelas de administración privada han tendido a responder seleccionando estudiantes con mayores probabilidades de éxito antes que con mejoras en la calidad. También se achaca a la libre elección de escuela el incremento en la brecha entre los pobres y los más ricos, la desmoralización de los "perdedores", la "enseñanza para las pruebas" y la reducción del currículo, así como la generación de algunas prácticas corruptas.

Mejorar la gestión pedagógica e institucional en las escuelas o áreas escolares

Uso de los resultados para diseñar planes de mejoramiento escolar

Es parte de un discurso común asumir que las evaluaciones a gran escala tendrán un impacto en los logros de aprendizaje solamente si sirven para generar cambios positivos en la escuela y el aula, actuando como insumo para el diseño, implementación y evaluación de estrategias y planes de mejoramiento escolar. La descentralización educativa o la autonomía escolar que se viene impulsando en varios países de la región han traído consigo una serie de normas e incentivos para el diseño de planes de mejoramiento educativo a nivel subnacional, local e institucional. Estos proyectos de desarrollo institucional, planes de mejora, proyectos de innovación, etc., deben basarse, al menos parcialmente, en una lectura de los resultados de las evaluaciones.

Donde las pruebas son censales y los resultados son devueltos a cada escuela –con guías o capacitación sobre cómo interpretarlos, además de con asistencia técnica sistemática en el proceso–, la posibilidad de diseñar planes consistentes, centrados en mejoras pedagógicas y en el mejoramiento de los logros se incrementa considerablemente. También ocurre esto cuando se asocia un incentivo económico a la elaboración de buenos proyectos, lo que permitiría su implementación. Está pendiente aún un análisis sistemático de la consistencia de diagnósticos y planes de acción y de la viabilidad de estos planes, tales como se vienen elaborando en Colombia, Paraguay, Perú, Chile o Uruguay, entre otros. Está también pendiente una evaluación de su impacto sobre el mejoramiento de los aprendizajes.

Uso de los resultados para proveer oportunidades de desarrollo profesional para profesores y supervisores

Según los especialistas educativos, el uso de los resultados de las pruebas para la formación y capacitación docente es tan importante como su uso para el planeamiento escolar.

En ese sentido, una de las primeras y más sencillas formas de utilización de los resultados de las pruebas en América Latina ha sido la publicación y distribución más o menos extensa de folletos de "análisis de ítems". La publicación de los resultados generales de las pruebas ha estado acompañada por la publicación de ítems seleccionados de las pruebas para cada una de las

áreas curriculares, usualmente reflejando distintos niveles de dificultad. Se suele presentar el ítem completo con las múltiples opciones de respuesta, seguidas normalmente por una descripción de las habilidades curriculares específicas que estaban siendo evaluadas con esa pregunta y una explicación de por qué se les consideraba lo suficientemente importantes como para ser evaluadas en la prueba. Las publicaciones suelen analizar también las razones probables por las que los estudiantes podrían elegir la respuesta correcta o alguna de las respuestas incorrectas y, a veces, ofrecen algunas sugerencias –tanto conceptuales como didácticas– sobre cómo los profesores podrían ayudar a los estudiantes a desarrollar la habilidad prevista.

En países donde las pruebas solo se aplicaron a muestras de estudiantes de algunos grados, estas publicaciones con frecuencia sirvieron como una primera oportunidad para que los profesores y supervisores se informaran sobre los propósitos y métodos de las evaluaciones nacionales. Por esto tales informes "pedagógicos" generalmente incluyen algunas páginas introduciendo a los educadores al significado y uso potencial de las evaluaciones para su propio desarrollo profesional y al análisis de su posible desempeño en este tipo de pruebas.

Asimismo, y dado que los informes generales rara vez logran llegar hasta las escuelas, las publicaciones incluyen también un resumen de los resultados nacionales o subnacionales. En la mayoría de países se esperaba que no solo todas las escuelas recibieran copias de estos materiales, sino que estas y los distritos escolares tomarían la iniciativa para analizarlos. Sin embargo, en muchos casos la distribución del material estuvo seriamente limitada por falta de presupuesto y, en la mayoría de países, no se ha hecho ningún esfuerzo para supervisar de qué modo se está utilizando estas publicaciones en la práctica, quiénes las están usando, ni mucho menos con qué resultados.

En América Latina, una vez que las evaluaciones comenzaron a estar –parcial o totalmente– guiadas por marcos referidos a criterios, los informes dirigidos a los docentes se han centrado en descripciones de los distintos niveles de desempeño en cada una de las áreas o criterios de aprendizaje, cada una con ejemplos de uno o varios ítems o tareas que los estudiantes que alcanzan ese nivel de desempeño podrían responder o realizar correctamente.

Esto ha ocurrido en Argentina, Brasil, Perú y Chile, entre otros. Sin embargo, seguimientos realizados sobre el uso de estos instrumentos sugieren que solo las mejores escuelas reportan un mayor uso, mientras que aquellas que más mejoras necesitan no están en capacidad o no están dispuestas a utilizar los resultados de las pruebas para este fin. Actualmente se está impulsando a las escuelas a organizar eventos especiales con el personal con el objetivo específico de discutir los informes y existe una convicción de que para ello es indispensable proveer asistencia técnica a las escuelas.

A pesar de haber realizado consultas con varias fuentes, no se ha encontrado ninguna información sistemática sobre maneras específicas en las cuales los resultados de las pruebas se estén utilizando para diseñar capacitaciones o programas de desarrollo profesional. De hecho es muy posible que esté habiendo mucha actividad en este sentido, dado que los funcionarios ministeriales suelen contar la capacitación de profesores entre los usos dados a los resultados de las pruebas.

Sin embargo, hay razones para creer que el contenido de muchas de tales actividades bien puede consistir solo en un análisis superficial de los resultados de las pruebas nacionales y en una introducción general para los profesores sobre el sistema de evaluación. La identificación y descripción de formas más promisorias en que los resultados de las pruebas puedan ser utilizados como

material para la capacitación docente es aún una tarea pendiente. En su capítulo en este libro, Pedro Ravela analiza en detalle la evaluación y el desarrollo profesional de los docentes.

CONCLUSIONES

Una primera conclusión de esta rápida revisión es que los resultados de las pruebas parecen estar siendo utilizados más de lo que usualmente se sospecha. Dada la existencia relativamente corta de los sistemas de evaluación de la región y la inestabilidad del compromiso político con su desarrollo, no sorprende que distintos países utilicen los resultados en menor medida de lo que se esperaría. Sin embargo, si se toma en cuenta la grave escasez de recursos para el mejoramiento educativo y la existencia de muchos programas valiosos que valdría la pena financiar, parecería necesario redoblar los esfuerzos para dar un uso adecuado a los resultados de las evaluaciones.

Aunque la mayoría de países latinoamericanos no iniciaron el desarrollo de sus sistemas de evaluación con ideas claras sobre los usos que se podría dar a los resultados, una vez que empezaron a recoger información fueron encontrando diversas maneras de usarla. Así, por ejemplo, la información ha sido utilizada para: el diseño e implementación de políticas; la asignación de recursos; estimular cambios pedagógicos y promover la responsabilización entre varios actores e instituciones educativas. Ha habido además una considerable evolución en la forma de presentar los resultados a diversas audiencias, y cada vez se está dando mayor atención a la necesidad de incluir información sobre factores de contexto que deben ser tomados en cuenta en el momento de interpretar los resultados.

Dado que en muchos lugares la prensa se ha convertido en el principal comunicador de los resultados a profesores, padres y tomadores de decisiones, es claro que no se han realizado esfuerzos suficientes para educar a los medios de comunicación de modo que estos se conviertan en aliados de los esfuerzos para movilizar al público en apoyo de la educación, en vez de actuar, conscientemente o no, como enemigos del sistema público. Esto requerirá que la mayoría de agencias de evaluación asigne una cantidad considerablemente mayor de recursos que los que se han estado invirtiendo en esfuerzos de difusión (tiempo del personal para definir claramente los principales mensajes que deben ser transmitidos al público en general a través de la prensa; tiempo para responder a interpretaciones incorrectas o no deseables de los resultados de las pruebas; así como más dinero).

Aunque ahora está empezando, el efecto de los resultados de las pruebas en la renovación curricular y el establecimiento de estándares también parece prometedor –cabe notar que muchos países pusieron en marcha sus sistemas de evaluación al mismo tiempo que daban inicio a reformas curriculares cuya llegada a las aulas aún tenía que ser asegurada–. Sin embargo, el diseño de las pruebas aún necesita mejorarse para permitir investigaciones adicionales de las implicancias pedagógicas de sus resultados. La posibilidad de utilizar los resultados para mejorar la planificación en el ámbito de las escuelas también está emergiendo como una línea de acción interesante y requerirá de aplicaciones de tipo censal o de la provisión de instrumentos para la autoevaluación y el análisis colectivo de los resultados.

Otra manera de que las evaluaciones pueda contribuir a mejorar el desempeño docente y los logros de aprendizaje es la producción de materiales para el desarrollo profesional que estén basados en análisis detallados de los resultados de las pruebas y la generación de espacios institucionales y eventos de capacitación en los cuales estos puedan ser apropiados y usados a cabalidad.

El análisis de ítems seleccionados de las pruebas que es difundido por las agencias de evaluación ha facilitado con frecuencia entre los profesores la reflexión sobre sus propias expectativas para el aprendizaje de sus estudiantes, así como la reconsideración por parte de los especialistas de cuán adecuados son los objetivos de aprendizaje que han definido para varios grados. El acceso directo a los instrumentos de evaluación también parece una poderosa manera de aliviar los temores y prejuicios existentes sobre la gama supuestamente limitada de habilidades que las pruebas pueden evaluar. Hoy día, incluso los ítems de opción múltiple pueden explorar procesos de razonamiento complejos y pueden servir para identificar fuentes de error ligadas a las prácticas de aula, si bien no debe aspirarse a sustituir el rol de las evaluaciones formativas en aula, otra área crítica en la gestión del aprendizaje.

Sin embargo, si se quiere asegurar el éxito de las estrategias de difusión y uso de los resultados, se tendrá que prestar mayor atención a las culturas organizacionales, tanto a nivel de las escuelas como en los niveles intermedios de la administración educativa. Las directrices tradicionales de tipo vertical (*top-down)* con frecuencia han fallado en tener los resultados anticipados cuando los intereses burocráticos y profesionales aparecen amenazados por la nueva información disponible, o cuando simplemente no existen las capacidades para ser movilizadas.

Los resultados de las evaluaciones y su papel en las políticas educativas

Álvaro Marchesi

DIFERENTES PERSPECTIVAS Y CONTROVERSIAS EN LA EVALUACIÓN EDUCATIVA

Como se ha podido comprobar en la lectura de los diferentes capítulos del libro, hay múltiples desajustes y discrepancias en el ámbito de la evaluación como sucede en casi todos los temas educativos relevantes. Los tres más importantes se refieren a la coherencia entre el significado de la educación y los objetivos de las evaluaciones, a la definición y las finalidades de las evaluaciones y a las relaciones entre la evaluación, la política y el cambio educativo.

Significado de la educación y objetivos de la evaluación

Las funciones de la educación escolar se han ido ampliando en la medida en que se incrementaban las exigencias sociales. Ya no se trata solo de aprender determinados conocimientos, sino que es preciso "aprender a conocer, aprender a hacer, aprender a vivir juntos y aprender a ser", de acuerdo con la acertada propuesta del informe Délors. Y para ello es preciso que los profesores adquieran las competencias necesarias y que las escuelas se organicen de tal manera que estos objetivos formen parte indisociable del trabajo educativo.

Pero no solo se han ampliado los objetivos de la educación. También se tiene conciencia de que el logro de dichos objetivos no puede atribuirse de forma exclusiva a la responsabilidad del sistema educativo –administraciones educativas, maestros y alumnos–, sino que debe recaer también en otras administraciones públicas y sectores sociales: municipios, familias, centros de salud, medios de comunicación, voluntarios…; y que si importante es el tiempo escolar, tampoco deben olvidarse las horas que el alumnado pasa fuera de la escuela y las actividades que realiza en ellas.

Lo que aparece en muchas ocasiones –y en ello se encuentra la primera de las tensiones– es la habitual distancia entre el discurso pedagógico-educativo y el discurso específico de los programas de evaluación. Mientras que en los primeros, como se acaba de apuntar, aparecen con fuerza elementos vinculados a la adquisición de conocimientos, al desarrollo social y a la formación en actitudes y valores, así como referidos a la participación activa de diferentes agentes e instituciones sociales, el ámbito de la evaluación escolar suele quedar restringido a la organización de las escuelas, a la práctica docente de los maestros en sus disciplinas específicas y principalmente a los aprendizajes de los alumnos en lengua, matemáticas y conocimiento científico.

Diferentes finalidades

Existe un acuerdo inicial entre la gran mayoría de los estudiosos del tema sobre el significado de la evaluación: obtener información a través del conocimiento de la realidad y de procedimientos

científicos contrastados. A partir de ahí aparecen las diferencias más notables. Un primer tipo de propuestas son las que ponen el acento principal en la valoración del proyecto: evaluar es determinar el mérito o el valor de algo (Scriven, 1991, p. 139). Otro conjunto de orientaciones insiste más bien en la adopción de decisiones y en el cambio de los programas como objeto de la evaluación. No se trata, afirmará Cronbach (1982), de ser un árbitro que establezca lo bueno y lo malo, sino que la tarea principal del evaluador es ayudar a adoptar decisiones que contribuyan a la mejora del objeto evaluado. La evaluación, en cierta medida, deberá acompañar al desarrollo de la educación e integrarse en ella. Finalmente, existe un tercer grupo de aproximaciones a la función de la evaluación referidas al desarrollo del conocimiento y a la sensibilización de los sectores interesados en la educación. En este caso, el acento se sitúa en la participación, en el diálogo y en el cambio educativo, por lo que el evaluador sería más bien un mediador en el proceso de construcción de los conocimientos apropiados y un agente indirecto en la mejora de la educación.

En todas estas finalidades, especialmente en aquella referida a la valoración del mérito, está presente de una forma o de otra un determinado modelo educativo y una opción de valor, lo que también está presente en la acción política. Por ello, las relaciones entre la evaluación y la política educativa pueden comprenderse mejor una vez que se hayan comentado las tres finalidades principales de la evaluación: información, valoración y orientación para la adopción de decisiones. Al final, en el último apartado, se discutirán de forma específica las relaciones entre evaluación, política y cambio educativo, que además constituyen el tercero de los desajustes al que se ha hecho mención al inicio del capítulo.

EL MOSAICO INFORMATIVO

Selección de la información

La selección de la información y su presentación son dos momentos clave del proceso de evaluación. En ellos se encuentran gran parte de los dilemas a los que debe enfrentarse la evaluación, y de su orientación va a depender la forma en la que se desarrolle la secuencia posterior.

La primera cuestión se refiere a la selección de la información y afecta fundamentalmente a las dimensiones que van a ser objeto de la evaluación y a la obtención de los datos que se consideran relevantes. No cabe duda de que se evalúa lo que se valora, por lo que el enfoque de la evaluación, en el que se refleja un determinado modelo educativo, tiene una importancia indudable.

Elegir la evaluación de las competencias en matemáticas y en lengua frente a otras opciones como educación artística, convivencia, integración social o ciudadanía, e incluir o no el contexto social y cultural, los recursos de las escuelas o la formación de los docentes son alternativas que sin duda van a tener repercusiones en la importancia que se otorga a determinados aprendizajes escolares y en la interpretación de los resultados obtenidos. Por ello es importante una cierta planificación de las evaluaciones para que todas o casi todas las dimensiones relevantes en la acción educativa puedan ser tenidas en cuenta.

Presentación y utilización de la información

La segunda cuestión afecta a la presentación y utilización de la información y, dentro de ella, al tipo de comparación que contribuye a comprender el alcance de la información obtenida. Conviene señalar, en primer lugar, que algún tipo de comparación con algún elemento de referencia

es necesario para comprender de forma más precisa el significado de la información obtenida y para realizar la valoración del objeto de la evaluación. También merece la pena recordar que la forma de comparación seleccionada no condiciona su modo de presentación, bien sea abierto a todo el público, limitado a determinados sectores o reservado.

Alejandro Tiana, en un artículo que formó parte del Programa de Evaluación de la Calidad de la Educación gestionado por la propia OEI hace más de una década (1996), apuntó tres tipos principales de comparaciones, con sus ventajas e inconvenientes. El primero supone la selección de un criterio o indicador que sirve de referente para realizar la valoración oportuna. El segundo elige la comparación entre las distintas entidades que participan en la evaluación, bien sean países, escuelas o alumnos, y determina la posición de cada una en relación con el resto. El tercer tipo intenta evitar los riesgos de las anteriores al tener que seleccionar un criterio o comparar realidades educativas que actúan en muy diversas circunstancias, y opta por la comparación del mismo objeto educativo a lo largo del tiempo.

El modo de comparación elegido manifiesta un determinado modelo educativo sobre la calidad de la enseñanza e incluso una estrategia propia para mejorar la educación, lo que sucede también en la forma como se presentan los resultados de la evaluación. La publicación ordenada de los resultados académicos de los alumnos de las escuelas de una zona o de un país, por ejemplo, suele implicar la asociación de esos resultados con la calidad de cada escuela y la importancia otorgada al esfuerzo autónomo que cada una de ellas debe desarrollar. También se presupone que la elección por parte de las familias de las mejores escuelas para sus hijos va a ser el mecanismo o revulsivo para que todas las escuelas se esfuercen por mejorar sus resultados y conseguir mayor número de alumnos y posiblemente con ellos también mayores recursos económicos.

A la hora de presentar la información hay que tener en cuenta los riesgos a los que se enfrenta. Uno de ellos es la simplicidad en la transmisión de los resultados a la opinión pública, que en muchas ocasiones no se corresponde con el rigor de los informes, lo que provoca normalmente una comprensión inadecuada de la compleja realidad educativa de un país, de una escuela o de un programa educativo. Las últimas naciones en una comparación internacional de resultados o las últimas escuelas en una evaluación de un país o región pueden quedar marcadas por esa posición sin que se repare en las limitaciones de su contexto sociocultural, en la composición de su alumnado o en el posible esfuerzo realizado en los últimos años. En su capítulo en este libro, Patricia McLauchlan de Arregui ha analizado en detalle el tema de la difusión de los resultados de la evaluación.

Un segundo riesgo apunta al uso indebido de inferencias no sostenidas por la evaluación realizada, lo que en muchas ocasiones se produce desde las evaluaciones individuales, sobre todo de los logros académicos de los alumnos, hacia el funcionamiento y la calidad de las escuelas y del conjunto del sistema educativo. A veces, sin embargo, dichas inferencias están de alguna manera implícitas en la evaluación correspondiente o en los modelos educativos subyacentes para los que el rendimiento académico de los alumnos es la genuina expresión de la calidad de una escuela o del sistema educativo.

Por todo ello es preciso abordar las evaluaciones y la utilización de la información con el máximo cuidado, teniendo en cuenta no solo los efectos conceptuales e instrumentales del conocimiento y de la información generada, sino también los políticos y los simbólicos. Los resultados de la evaluación pueden convertirse en un poderoso instrumento de cambio si se cuida su utilización, si se evitan los riesgos de su mal uso e incluso de su falta de uso, si se sientan las bases para una

correcta interpretación de los resultados y se establece un marco de participación y de debate que permita el lógico contraste de opiniones diversas ante el pluralismo valorativo presente en la evaluación.

EL MÉRITO Y LA INTERPRETACIÓN EN LA EVALUACIÓN

Existe una amplia coincidencia entre los profesionales que trabajan en el campo de la evaluación educativa en que la determinación del mérito o del valor del objeto evaluado es uno de los rasgos distintivos de los procesos de evaluación. A partir de la información obtenida por medio de los procedimientos técnicos adecuados, es preciso valorar cómo está funcionando aquello que se evalúa, si cumple los objetivos previstos, sus logros y sus carencias principales y, en la medida de lo posible, las causas o, si esto no fuera posible, los correlatos de esta situación. La determinación del mérito supone un juicio de valor en el que entran en juego factores educativos, políticos y éticos.

Como ya se ha señalado anteriormente, los modelos educativos y las opciones de valor operan también en las primeras fases del proceso de evaluación y las decisiones adoptadas en ellas influyen en las fases ulteriores y definitivas. También la elección del modo de comparación de los resultados influye en la valoración del objeto estudiado y es a su vez expresión de una determinada concepción educativa. La comparación ordenada de las escuelas en función de los resultados directos de los alumnos no permite la misma valoración que si se tiene en cuenta el valor agregado que las escuelas aportan a los logros de los alumnos en función de su nivel inicial y de su contexto sociocultural. La valoración de una escuela tampoco sería la misma si se comparan sus resultados con otras escuelas o si la comparación se realiza con ella misma en diferentes períodos de tiempo.

La decisión sobre el mérito que se atribuye a las instituciones evaluadas –lo mismo se podría decir de los sistemas y de los individuos– es, como se acaba de afirmar, la expresión de una opción de valor. En consecuencia, las decisiones que se asumen en la evaluación, como en la educación en su conjunto, no son neutras, sino que reflejan una determinada concepción del papel de la educación en la sociedad, de los objetivos educativos que deben destacarse y de los modelos de cambio que se consideran más pertinentes. No es extraño por ello que sea en el campo de la valoración del mérito de la realidad evaluada en donde las relaciones entre la evaluación y la política se hacen presentes de forma más nítida.

En la determinación del mérito o valor del objeto evaluado ocupa un papel relevante la interpretación o explicación que se realiza de la información obtenida. La toma en consideración de unos factores u otros y la metodología utilizada orienta en una u en otra dirección el discurso interpretativo. Además, la posibilidad de incorporar informaciones procedentes de otras evaluaciones puede ampliar la realidad evaluada. Desde esta perspectiva adquiere una especial relevancia la integración de las informaciones obtenidas en las diferentes evaluaciones. Hay que reconocer, sin embargo, que la pluralidad de modelos, opciones de valor y metodologías utilizadas complican enormemente esta labor.

Pero también la tarea interpretativa, en gran medida también valorativa, tiene una especial relación con otra de las finalidades de la evaluación: la información para la adopción de decisiones y para el cambio. En función de las relaciones establecidas entre las variables seleccionadas, el énfasis de la orientación puede estar en el mérito de la realidad educativa evaluada, en las barreras

que determinadas condiciones imponen al objeto evaluado o en la búsqueda de nuevas informaciones que permitan comprenderlo mejor. Además, el énfasis en tales o cuales dimensiones que contribuyen a la interpretación de la realidad educativa no solo puede orientar las decisiones, sino que sitúa el foco de la responsabilidad en determinadas instituciones, sectores o personas. Por ello sería preciso que la evaluación incluyera de alguna manera a todos aquellos que pudieran tener alguna responsabilidad con el desempeño del objeto evaluado.

Un ejemplo puede ayudar a clarificar este planteamiento. Supongamos que se ha realizado una evaluación de un programa de integración de alumnos con necesidades educativas especiales en la escuela ordinaria y se ha comprobado que estos alumnos tienen más dificultades de relación social con sus compañeros que aquellos otros que estudian en una escuela especial. ¿Cómo interpretar estos datos y orientar a partir de ellos? ¿Sería correcto sugerir que no es positiva, al menos en este ámbito, la inclusión educativa? ¿O solo que hay que cambiar determinadas condiciones –formación de maestros, recursos humanos y materiales, actitudes de los maestros, de las familias y de los alumnos, actividades de ocio que desarrollan los municipios, etc.– para mejorar el proyecto? ¿Qué condiciones son las más importantes, de quién es la responsabilidad principal y cómo se ha constatado en esta evaluación? ¿O se podría pensar que esos resultados se invertirían cuando se evaluara la adaptación social de los alumnos en contextos no escolares, lo que apuntaría a otro tipo de evaluación en un nuevo contexto?

EVALUACIÓN, ADOPCIÓN DE DECISIONES Y CAMBIO EDUCATIVO

Se acaba de señalar la estrecha relación entre el juicio de valor sobre una realidad educativa y la orientación para la posible adopción de decisiones. En el imaginario de los evaluadores opera la creencia de que sería deseable que las informaciones que proceden de evaluaciones rigurosas sirvan como referente principal en las decisiones de los políticos para orientar los proyectos educativos.

Sin embargo, como fácilmente puede constatarse por la experiencia de evaluadores y de políticos, la situación es bien distinta. La razón fundamental no es tanto el rechazo de los resultados de las evaluaciones, sino el hecho de que las decisiones políticas han de tener en cuenta diferentes fuentes de información, necesitan tomar en consideración las demandas, justificadas o no, de los actores concernidos por la decisión, aconsejan negociación y consenso en muchas ocasiones, han de tener en cuenta objetivos políticos generales y no deben olvidar las repercusiones sociales de la decisión. La valoración del momento y de la oportunidad, así como la necesidad de aunar fuerzas, son condicionantes importantes de las decisiones políticas. En este juego de fuerzas en las que se mueven las decisiones políticas, lo que hoy parece acertado porque resuelve un problema inmediato, mañana se vuelve equivocado porque ha abierto otros muchos problemas. No es extraño por todo ello que no pocos evaluadores consideren que los resultados de las evaluaciones van por un lado y las decisiones políticas por otro, y que los políticos tienden en demasía a utilizar la información proporcionada por las evaluaciones al servicio de sus intereses inmediatos.

Las limitaciones que se acaban de apuntar de la capacidad de las evaluaciones para influir en las decisiones políticas no deben ser motivo de desánimo. Sería una concepción demasiado estrecha considerar que la evaluación solo puede intervenir en la mejora de la educación a través de su influencia en el ámbito de las decisiones políticas. Los resultados de las evaluaciones influyen en la realidad educativa por muy diversas vías, entre las que sin duda es preciso destacar la referida a las decisiones en el ámbito de la política educativa. Pero hay otras muchas formas en la que se constata

su influencia en el cambio educativo: en su impacto en el funcionamiento de las instituciones, en su efecto en las actitudes, creencias y comportamientos de los docentes, en el mayor compromiso de determinadas instituciones como la familia o los municipios y en el tipo de mensajes que transmiten los medios de comunicación. Todas estas instituciones, sectores y personas pueden a su vez influir en los tomadores de decisión, por lo que no se debe olvidar que la información que procede de las evaluaciones se mueve por caminos diversos e inesperados.

La política educativa y la evaluación pueden establecer relaciones recíprocas, pero ambas pueden a su vez influir en otros agentes e instituciones educativas y en la propia actividad educadora. Este tipo de efectos puede realizarse de forma coordinada o de forma independiente por falta de comunicación o de entendimiento. Lo que determina el impacto real de la evaluación por una vía o por otra es la consistencia y relevancia de la información obtenida, así como la forma y persistencia de transmitir la información a los distintos sectores educativos. Al final, lo importante no es tanto la manera como se utiliza la información, sino el impacto que provoca en los sectores implicados.

Desde esta perspectiva, los evaluadores deberían comprender cómo trabajan los responsables políticos, cuáles son sus limitaciones y cómo adoptan decisiones. Pero también sería necesario que tuvieran en cuenta qué tipo de instituciones y agentes tienen mayor incidencia en el cambio educativo, cómo se llega a ellos y qué barreras han de superarse para que los resultados de las evaluaciones tengan efecto en el sistema educativo. Los modelos sobre la utilización de la evaluación han de ampliarse al análisis de su impacto y este debería incorporar también una teoría sobre el cambio educativo.

Estas razones avalan la importancia de conocer las características del receptor de la información, así como el papel de los intermediarios de la misma: supervisores, directores, asesores y líderes de opinión en campo social y educativo. La atención a los medios de comunicación es primordial para conseguir el efecto deseado, por lo que debería cuidarse la precisión del mensaje que se quiere transmitir, algo que suele ser contrario al estilo habitual de los titulares periodísticos.

La importancia otorgada a la influencia de las evaluaciones en muy diversas formas en los distintos ámbitos del sistema educativo, desde los responsables políticos hasta los profesores, pone de relieve una nueva función de la evaluación que, sin olvidar las ya mencionadas, debería tenerse en cuenta: la iluminación de la realidad educativa con el fin de facilitar su compresión e interpretación, sugerir orientaciones a los diversos sectores implicados en función de su responsabilidad y favorecer su participación para garantizar un mayor compromiso con sus resultados. Este planteamiento puede facilitar las relaciones en ocasiones difíciles entre la evaluación y la política educativa[1].

EVALUACIÓN Y POLÍTICA EDUCATIVA

Diferentes relaciones

A lo largo de las páginas anteriores se ha ido apuntando la continua relación entre la evaluación y las decisiones políticas. Los programas y proyectos de reformas que han de ser evaluados pro-

[1] En el capítulo de Marta Kisilevsky incluido en este libro se analiza otra de las relaciones entre evaluación y política educativa, la que se refiere a valorar los programas y actuaciones políticas.

ceden de decisiones políticas, los juicios de valor que se reflejan en los informes de evaluación contactan con los juicios de valor presentes en la arena política y las orientaciones que se derivan de las evaluaciones se incorporan también al debate político. A pesar de los esfuerzos de los evaluadores de mantener la máxima neutralidad posible, es necesario ser conscientes de las permanentes implicaciones políticas que las evaluaciones conllevan.

En un sugerente trabajo sobre las relaciones entre evaluación y política, Karlsson y Konner (2005) describen tres perspectivas sobre dichas relaciones: la primera posición sostiene que es posible y deseable separar evaluación y política; la segunda, que es posible separar evaluación y política cuando se proporciona información, pero que no lo es cuando se proporcionan juicios de valor y que además no es conceptualmente deseable; la tercera posición mantiene que no es ni posible ni deseable, operacional y conceptualmente, separar la evaluación de la esfera de la política.

No corresponde a este texto analizar cada uno de estos enfoques, aunque a lo largo de estas páginas se ha tratado de argumentar a favor de una posición que se aproxima a la tercera de ellas. De lo expuesto en estas páginas se deduce que la concepción tradicional del evaluador como científico objetivo y neutral que se enfrenta a la realidad con sus propias herramientas técnicas y metodológicas, sin contaminarse con las disputas políticas que le rodean, es difícilmente sostenible. El evaluador está inmerso en el debate político que afecta a las reformas y a los programas que son objeto de su trabajo y ha de ser consciente de los diferentes intereses que están en juego e incluso que él mismo, con sus propios valores, forma parte de esos intereses. Como señalan Karlsson y Konner, la mejor imagen de un evaluador es la de un profesional bien entrenado, que trabaja en un contexto con implicaciones implícitas o explícitas políticas, culturales y personales que influyen en todas las fases del proceso de evaluación, y que mantiene una actitud respetuosa con las diversas posiciones en juego.

La autonomía de la evaluación

Las relaciones entre evaluación y política son en múltiples ocasiones asimétricas, sobre todo cuando los evaluadores dependen del poder político o cuando existe el riesgo de que determinadas valoraciones puedan erosionar la confianza política y provocar la marginación en futuras evaluaciones. Por ello no es fácil que el evaluador mantenga su independencia de criterio. Su mayor o menor autonomía dependerá de su propia capacidad, pero también de la forma de trabajo de los poderes públicos. En función del tipo de interacción que se establezca entre ambos se podrán abrir o cerrar los caminos de una cooperación fructífera.

Entre las capacidades de los evaluadores está la consistencia teórica y técnica con la que presenten sus proyectos, su visión para situarse en la arena política, su habilidad para negociar en las diferentes fases del proceso de evaluación y su sensibilidad ante la diversidad de interpretaciones posibles, pues los estudios educativos difícilmente presentan conclusiones incontestables.

Por su parte, la forma de trabajo de los poderes públicos está en función de su grado de respeto a la autonomía de los proyectos de evaluación, de su nivel de aceptación de la información crítica a sus proyectos y de su sensibilidad para tenerla en cuenta a la hora de adoptar decisiones, de su capacidad de impulsar el debate y la participación, del reconocimiento de que el apoyo y la formación a los equipos responsables de la evaluación es una buena inversión, y de la voluntad de asegurar la coordinación de las evaluaciones para mejorar su interpretación y ofrecer una visión progresivamente integradora de la diversa realidad educativa.

Participación y diálogo

Por todo ello, la comprensión de las diferencias entre la esfera de la política y la esfera de la evaluación, así como de sus continuas interacciones, junto con una actitud que fomente el diálogo y la participación, contribuyen a garantizar el logro de las finalidades de la evaluación, especialmente iluminar la realidad educativa y colaborar en su transformación. La comunicación ha de mantenerse entre los poderes públicos y los evaluadores, pero también entre ellos y los diferentes agentes e instituciones educativas. El apoyo a redes diversas en las que se debatan los resultados de las evaluaciones, en las que se recojan propuestas y sugerencias y en las que sea posible la crítica y las posiciones alternativas supone un respaldo al proceso de evaluación y una forma de contribuir a su utilización y a su impacto.

La posibilidad de comunicación y de discusión enriquecedora se refuerza cuando se acepta la existencia del pluralismo en el seno de las evaluaciones, de manera similar a como está presente en la realidad social, cultural, educativa y política. Los proyectos de evaluación pueden orientarse desde diferentes opciones ideológicas, científicas y técnicas y conectar en mayor o menor medida con las plurales opciones políticas. La conciencia de esta situación ayuda a un diálogo enriquecedor. Esta defensa del pluralismo no conduce a situar en el mismo nivel a todas las evaluaciones. Aquellas que se orientan por los valores de calidad, equidad, participación, inclusión educativa y cohesión social son, a juicio del autor del presente texto, las que mejor contribuyen al logro de sus finalidades.

El evaluador debe ser consciente de que su trabajo está inmerso en debates científicos a los que debe hacer frente, pero que también está concernido con cuestiones políticas y éticas de las que no puede evadirse. La respuesta coherente a estas demandas fortalece su figura profesional y amplía el impacto educativo y social de su tarea.

Bibliografía

Aguilar Villanueva, L. (ed.) (2000), *Problemas públicos y agenda de gobierno*, México, Miguel Ángel Porrúa.

Ainscow, M. (2001), *Desarrollo de escuelas inclusivas, ideas, propuestas y experiencias para mejorar las instituciones escolares*, Madrid, Narcea Ediciones.

Archbald, D. y Newman, F. (1992), "Approaches to assessing academic achievement", en Tom, A. R. (ed.), *Toward a new science of educational testing*, Nueva York, State University of New York Press.

Ardoino, J. (2000), "Consideraciones teóricas sobre la evaluación en educación", en Rueda, M. y Díaz, F. (comp.), *Evaluación de la docencia. Perspectivas actuales*, pp. 23-37, México, Paidós.

Arnáiz, P. (2003), *Educación inclusiva. Una escuela para todos*, Madrid, Editorial Aljibe.

Arregui, P. (2001), *Sistemas de determinación y evaluación de metas de logros de aprendizaje escolar como instrumento para mejorar la calidad, la equidad y la responsabilización en los procesos educativos en América Latina. Análisis de prospectivas de la educación en América Latina y el Caribe*, Santiago de Chile, UNESCO.

Báez, B. (1994), "El movimiento de escuelas eficaces. Implicaciones para la innovación educativa", *Revista Iberoamericana de Educación*, 4, pp. 102-103.

Ballart, X. (1992), *¿Cómo evaluar programas y servicios públicos? Aproximación sistemática y estudios de caso*, Madrid, Ministerio para las Administraciones Públicas.

Bardach, E. (1996), "El problema de la definición de problemas en el análisis de políticas", en Aguilar Villanueva, L. (ed.), *Problemas públicos y agenda de gobierno*, pp. 219-234, México, Miguel Ángel Porrúa.

Barber, M. y Mourshed, M. (2007), *How the world's best-performing school systems come out on top*, McKinsey & Company, disponible en: http://www.mckinsey.com/clientservice/socialsector/resources/pdf/Worlds_School_Systems_Final.pdf. Traducción al castellano, PREAL (2008), *Cómo hicieron los sistemas educativos con mejor desempeño del mundo para alcanzar sus objetivos*, Santiago de Chile, PREAL, documento de trabajo n.° 41, disponible en: http://www.preal.org/publicacion.asp.

Benveniste, L. (2000), *La evaluación del rendimiento académico y la construcción de consensos en Uruguay*, Lima, GRADE-PREAL.

Bray, M., Adamson, B. y Mason, M. (2007), *Comparative education research. Approaches and methods*, Hong Kong, CERC-The University of Hong Kong.

Carnoy, M. (2004), *Las reformas educativas en la década de 1990. Un estudio comparado de Argentina, Chile y Uruguay*, Proyecto Alcance y resultados de las reformas educativas en Argentina, Chile y Uruguay, desarrollado por el Banco Interamericano de Desarrollo, los Ministerios de Educación de Argentina, Chile y Uruguay y el Grupo Asesor de la Universidad de Stanford, Buenos Aires, Akian Gráfica Editora.

Carnoy, M. y Marshall, J. (2005), "Cuba's academic performance in comparative perspective", *Comparative Education Review*, vol. 49, 2.

Carr, W. (1995), *Una teoría para la educación. Hacia una investigación educativa crítica*, Madrid, Ediciones Morata.

Casassus, J. (2003), *La escuela y la desigualdad*, Santiago de Chile, LOM Ediciones.

Castillero, B., Díaz, J., Morales, F. y Rino, J. (2002), *Gestión y supervisión en el centro de educación básica*, Costa Rica, Coordinación Educativa y Cultural Centroamericana, n.º 36.

Center for Educational Testing and Evaluation (1995), *1995 Kansas science assessment. Examiner´s manual*, Lawrence, The University of Kansas Printing Service.

Center for Educational Testing and Evaluation (1995), *1995 Kansas science assessment. Grade 8*, Lawrence, The University of Kansas Printing Service.

Chiara, M. y Di Virgilio, M. (2008), "Conceptualizando la gestión social", en *Gestión de la política social, conceptos y herramientas*, UNGS.

Cicourel, A. (2000), "Algunas cuestiones de teoría y método", en Goffman, E. y otros (eds.), *Sociologías de la situación*, pp. 97-117, Madrid, La Piqueta.

Cohen, E. y Franco, R. (1988), *Evaluación de proyectos sociales*, Buenos Aires, ILPES/ONU-CIDES/OEA-GEL, Colección Estudios Políticos y Sociales, 1.ª ed.

Coleman, J. S., Campbell, E., Hobson, C., Mc Partland, J., Mood, A., Weinfeld, F. y York, R. (1966), *Equality of educational opportunity*, Washington D. C., Department of Health, Education & Welfare, Office of Education.

Coll, C. y Martín, E. (2006a), *Vigencia del debate curricular. Aprendizajes básicos, competencias y estándares*, pp. 67-92, PRELAC, 9.

Coll, C. y Martín, E. (2006b), "Evaluaciones del rendimiento escolar y procesos de cambio curricular en España entre 1990 y 2005. Análisis de un caso", en Dols (ed.), *Recherches sur les curriculums, développement et pilotage*, pp. 78-109.

Coll, C., Martín, E. y Onrubia, J. (2001), "La evaluación del aprendizaje escolar, dimensiones psicológicas, pedagógicas y sociales", en Coll, C., Palacios, J. y Marchesi, A., *Desarrollo psicológico y educación*, vol. II, *Psicología de la educación*, pp. 549-567, Madrid, Alianza Editorial.

Comisión Europea (1999), *Selection and use of indicators for monitoring and evaluation. Evaluating socio-economic programmes*, vol. 2, Luxemburgo, Office for Publications of the European Communities.

Comisión Europea (2004), *Competencias clave para un aprendizaje a lo largo de la vida. Un marco de referencia europeo*. Puesta en práctica del programa de trabajo "Educación y Formación 2010", Grupo de trabajo B "Competencias clave", Comisión Europea, Dirección General de Educación y Cultura, disponible en:
http://www.educastur.princast.es/info/calidad/indicadores/doc/comision_europea.pdf.

Comisión para el Desarrollo y Uso del SIMCE (2003), *Evaluación de aprendizajes para una educación de calidad*, Santiago de Chile, Ministerio de Educación.

Condemarín, M. y Medina, A. (2000), *Evaluación auténtica de los aprendizajes. Un medio para mejorar las competencias en el lenguaje y comunicación*, Santiago de Chile, Editorial Andrés Bello.

Cox, C. (2008), "Book review. Tough choices or tough times, seeing with one eye", *Journal of Educational Change*, 9, pp. 83-85.

Cronbach, L. J. (1982), *Designing evaluations of educational and social programs*, San Francisco, Jossey Bass.

Cueto, S. (ed.) (2005), *Uso e impacto de la información educativa en América Latina*, Santiago de Chile, PREAL.

De Macedo, L. (2002), "Situação-problema, forma e recurso de avaliação, desenvolvimento de competências e aprendizagem escolar", en Perrenoud, P. y Gather, T. (ed.), *As competências para ensinar no século xxi. A formação dos profesores e o désafio da avaliação*, Porto Alegre. Artmed.

De Landsheere, G. (1994), *Le pilotage des systèmes d'éducation*, Bruselas, De Boeck.

Delors, J. (dir.) (1996), *La educación encierra un tesoro*, Informe a la UNESCO de la Comisión Internacional sobre Educación para el Siglo xxi, París, Organización de las Naciones Unidas.

DINIECE (2007), Gluz, N. y Chiara, M. (coord.), *Estudio evaluativo Programa Integral de Igualdad Educativa*.

DINIECE (2008), Chiara, M., *Guía para la evaluación de programas en educación*, Área de Investigación y Evaluación, Buenos Aires.

Esquivel, J. M. (2000), *Materiales del taller sobre evaluación del desempeño*, ofrecido a diversas audiencias en la región centroamericana. San José, CECC.

Esquivel, V. (2006), *De la integración a la educación inclusiva. Educación para todos*, conferencia dictada en el Congreso anual de la ANDE.

Estefanía, J. L. y López, J. (2003), *Evaluación externa del centro y calidad educativa*, Madrid, Editorial CCS.

EURYDICE (2004), *Evaluación de los centros de enseñanza obligatoria en Europa*, Madrid, Ministerio de Educación y Ciencia.

Fernández Ballesteros, R. (1996), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*, Madrid, Ed. Síntesis.

Ferrer, G. (2004), *Las reformas curriculares de Perú, Colombia, Chile y Argentina, ¿quién responde por los resultados?* Lima, GRADE.

Ferrer, G. (2006), *Sistemas de evaluación de aprendizajes en América Latina. Balances y desafíos*, Lima, PREAL-GRADE.

Ferrer, G. (en elaboración), *Estudio comparado internacional sobre procesos de elaboración e implementación de estándares de currículum en América Latina*.

Ferrer, J. G. y Arregui, P. (2003), *Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación. Criterios para guiar futuras aplicaciones*, documento n.º 23, Santiago de Chile, PREAL.

Forster, M. y Valverde, G. (2003), *La experiencia internacional en sistemas de medición, estudio de caso*, Santiago de Chile, Comisión para el Desarrollo y Uso del SIMCE.

Fullan, M. (1993), *Change forces. Probing the depths of educational reform*, Londres y Bristol, The Falmer Press.

Fullan, M. (1999), *Change forces. The sequel*, Londres y Filadelfia, The Falmer Press.

Gaulupeau, Y. (1992), *La France à l'école*, París, Gallimard.

Gilmore, A. (2005). *The impact of PIRLS (2001) and TIMSS (2003) in low and middle-income countries*, Nueva Zelanda, IEA.

Goh, S. C., Cousins, B. y Elliot, C. (2006), "Organizational learning capacity, evaluative inquiry and readiness for change in schools, views and perceptions of educators", *Journal of Educational Change*, n.º 7, pp. 289-318.

Goodlad, J. I. y Anderson, R. H. (1987), *The nongraded elementary school*, reedición con una nueva introducción de la edición de 1963, revisada de la House, E. R. (1993), *Professional evaluation. Social impact and political consequences*, Newbury Park, Londres y NuevaDelhi, SAGE, original de 1959, Nueva Cork, Teachers College, Columbia University.

Hintze, S. (2001), "Reflexiones sobre el conflicto y la participación en la evaluación de políticas sociales", *Reforma y Democracia de CLAD*, 21, presentado como ponencia en el VI Congreso Internacional del CLAD sobre la Reforma del Estado y la Administración Pública, Buenos Aires, noviembre, disponible en: http:// www.top.org.ar/publicac.htm.

House, E. y Howe, K. (1999), *Values in evaluation and social research*, Thousand Oaks, Sage Publications

IEA (2007a), *International Association for the Evaluation of Educational Achievement. Annual Report*, Amsterdam.

IEA (2007b), *International Association for the Evaluation of Educational Achievement* (ICSS), *Assessment Framework*, Amsterdam.

PIRLS (2006), *International Report*, Boston College.

INEE (2006), *Plan Maestro de Desarrollo 2007-2014*, México, INEE.

INEE (2007), *Propuesta y experiencias para desarrollar un sistema nacional de indicadores educativos*, México, INEE.

Ingvarson, L., Meiers, M. y Beavis, A. (2005), "Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy", *Education Policy Analysis Archives*, 13 (10).

Karlsson, O. y Konner, R. E. (2005), "The relationship between evaluation and politics", en Shaw, I. F., Greene, J. C. y Mark, M. M. (eds.), *The sage handbook of evaluation*, Londres, Sage Publications.

Kellaghan, Th. y Greaney, V. (2001), *Using assessment to improve the quality of education*, París, UNESCO/International Institute for Educational Planning.

Lane, S. (1993), "The conceptual framework for the development of a mathematics performance assessment instrument", *Educational measurement. Issues and practice*, verano, pp. 11-28.

Lane, S., Liu, M., Ankeonemann, R. y Stone, C. (1996), "Generalizability and validity of a mathematics performance assessment", *Journal of Educational Measurement*, n.º 33, primavera, pp. 71-92.

LeMahieu, P., Gitomer, D. y Eresh, J. (1995), "Portfolios in large-scale assessment, difficult but not impossible", *Educational measurement. Issues and practice*, otoño, pp. 11-28.

Linn, R., Baker, E. y Betebenner, D. (2002), "Accountability systems. Implications of requirements of the no child left behind act of 2001", *Educational Researcher*, n.º 31, 6, pp. 3-16.

LLECE (2008), Valdés, H. (coord.), *Los aprendizajes de los estudiantes de América Latina y el Caribe*, primer reporte de los resultados del Segundo Estudio Regional Comparativo y Explicativo (SERCE), Chile, UNESCO/OREALC.

López, N. (2007), *Las nuevas leyes de educación en América Latina. Una lectura a la luz del panorama social y educativo de la región*, Buenos Aires, CLADE-IIPE/UNESCO.

Mabry, L. (1999), "Writing to the rubric", *Phi Delta Kappan*, mayo, pp. 673-679.

Marchesi, A. y Martín, E. (comps.) (2002), *Evaluación de la educación secundaria. Fotografía de una etapa polémica*, Madrid, Fundación Santa María.

Martín-Moreno, Q. (2007), *Organización y dirección de centros educativos innovadores. El centro educativo versátil*, Madrid, Mc Graw Hill.

Martínez Rizo, F. (2002), "Revisión de literatura", en Varios autores, *La calidad de la educación en México. Perspectivas, análisis y evaluación*, Anexo 1, pp. 305-454, México, Miguel Ángel Porrúa-SEP.

Martínez Rizo, F. (2005), *La evaluación educativa, ¿oportunidad o peligro?,* conferencia presentada en el VIII Congreso Nacional de Evaluación Educativa, Hermosillo, noviembre.

Martínez Rizo, F. (2005), "El diseño de sistemas de indicadores educativos, consideraciones teórico-metodológicas", *Cuadernos de Investigación*, n.° 14, México, Instituto Nacional de Evaluación de la Educación.

Martínez Rizo, F. (2006a), "Las primarias comunitarias y su desempeño. Consideraciones a partir del estudio comparativo 2000-2005 del INEE", *Cuadernos de Investigación*, n.° 23, México, INEE.

Martínez Rizo, F. (2006b), "Pruebas de selección y pruebas para evaluar escuelas. Nuevas consideraciones sobre su uso y la difusión de resultados", *Cuadernos de Investigación*, n.° 21, México, INEE.

Martínez Rizo, F. *et al.* (2007), *Propuestas y experiencias para desarrollar un sistema de indicadores educativos*, México, INEE.

Martínez Rizo, F. (2007a), *Hacia un sistema nacional de evaluación educativa*, Mecanograma.

Martínez Rizo, F. (2007b), *Propuesta para el sistema de evaluación de la educación media superior*, Mecanograma.

MEC (2006), *Ley Orgánica de Educación (LOE) y Ley Orgánica del Derecho a la Educación (LODE)*, Madrid, Secretaría General de Educación.

MEC, Instituto de Evaluación (2007a), *La lectura en Pisa 2000, 2003 y 2006. Marco y pruebas de la evaluación*, Madrid.

MEC, Instituto de Evaluación (2007b), *PIRLS 2006. Estudio internacional de progreso en comprensión lectora de la IEA*, Madrid.

MEC, Instituto de Evaluación (2007c), *PISA 2006. Programa para la Evaluación Internacional de Alumnos de la OCDE*, Informe español, Madrid.

Ministère de l'Éducation Nationale, Enseignement Superieur et Recherche. Délégation a la Communication (2006), *Le socle commun des connaissances et des compétences*, Francia, Direction Générale de l'Enseignement Scolaire.

Mizala, A. y Romaguera, P. (1998), *Desempeño escolar y elección de colegios. La experiencia chilena*, documento de trabajo n.º 36, Serie Economía, Santiago de Chile, Centro de Economía Aplicada.

Mizala, A., González, P., Guzmán, A. y Romaguera, P. (2000), *Los maestros en Chile, carreras e incentivos*, Research Network Working Paper R-403, Washington, Banco Interamericano de Desarrollo.

Mokate, K. (2000), *El monitoreo y la evaluación, herramientas indispensables de la gerencia social. Notas de clase*. INDES/BID, disponible en: http:// www.indes.iadb.org.

Morduchowicz, A. (2002), *Carreras, incentivos y estructuras salariales docentes*, documento de trabajo n.° 23, Santiago de Chile, PREAL.

Morduchowicz, A. (2006), *Los indicadores educativos y las dimensiones que la integran*, Buenos Aires, IIPE-UNESCO, disponible en: http://www.iipe-buenosaires.org.ar.

Morduchowicz, A. y Duro, L. (2007), *La inversión educativa en América Latina y el Caribe. Las demandas de financiamiento y asignación de recursos*, Buenos Aires, IIPE-UNESCO, disponible en: http://www.iipe-buenosaires.org.ar.

Muñoz Repiso, M., Murillo, J., Barrio, R., Brioso, M. J., Hernández, M. L. y Pérez Albo, M. J. (2001), *La mejora de la eficacia escolar*, España, Ministerio de Educación Cultura y Deporte, CIDE.

Myers, R., Martínez, J. F., González, P. y Flores, B. (2006), *Evaluación y acompañamiento del Programa Escuelas de Calidad-Nivel Preescolar 2003-2006*, México, Universidad Pedagógica Nacional y Secretaría de Educación Pública, Programa Escuelas de Calidad.

National PTA (2000), *Building successful partnerships. A guide for developing parent and familiy involvement programs*, National Educational Service.

Navarro, J. C. (2003), "La evaluación y las actitudes de los docentes frente a ella, dificultades y alternativas de política", en *Evaluar las evaluaciones. Una mirada política acerca de las evaluaciones de la calidad educativa*, pp. 153-155, Buenos Aires, IIPE-UNESCO.

Niremberg, O. (2007), "Modelos para la evaluación de políticas, insumos para una mejora de la gestión", en Chiara, M. y Di Virgilio, M. M., *Gestión de la política social, conceptos y herramientas*, UNGS.

Oakes, J. y Heckman, P. E. (1994), *Age grouping of students*, en Husen, T. y Postlethwaithe, T. S. N. (eds.), *The International Encyclopedia of Education*, vol. 1, pp. 227-230, Oxford-Nueva York, Pergamon.

OCDE (2006), *Schooling for tomorrow. Personalising education*, París, OECD.

OCDE (2005), *Definition and Selection of Competencies (DeSeCo), Theoretical and Conceptual Foundations*, Strategy Paper.

OCDE (2005), *The definition and selection of key competencies. Executive Summary*, disponible en: https://www.pisa.oecd.org/dataoecd/47/61/35070367.pdf.

OCDE (2008), *Informe PISA 2006, competencias científicas para el mundo del mañana*, Madrid, Santillana.

Odden, A. y Kelley, C., *Paying teachers for what they know and do. New and smarter compensation strategies to improve schools*, California, Corwin Press, Inc.

OEI (1994), *Evaluación de la calidad de la educación*, documento base, Buenos Aires, OEI.

OREALC/UNESCO (2007), *Educación de calidad para todos, un asunto de derechos humanos*, Santiago de Chile, II Reunión Intergubernamental del Proyecto Regional de Educación para América Latina y el Caribe EPT/PRELAC.

Paulson, L., Paulson, P. y Meyer, C. (1991), "What makes a portfolio a portfolio?" *Educational Leadership*, n.º 49, mayo, pp.14-22.

Perrenoud, P. (1999), *Ensinar, agir na urgência, decidir na incerteza*, São Paulo, Artmed.

Poggi, M. (en prensa), *Old and new perspectives for traditional issues in planning*, París, IIEP-UNESCO.

Pont, B., Nusche, D. y Moorman, H. (2008), *Improving school leadership. Policy and practice*, París, OECD Publishing.

Popham, W. J. (2004), "Curriculum, instruction, and assessment, amiable allies or phony friends?", *Teachers College Record*, 106 (3), pp. 417-428.

PRIE (2007), *Balance de las misiones de diagnóstico de los sistemas de información educativa*, resumen ejecutivo, UNESCO y Secretaría de Educación Pública de México, disponible en: http://www.unesco.org/santiago.

Ravela, P. (2001), *¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?* PREAL.

Ravela, P. (ed.). (2001), *Los próximos pasos, ¿hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?* PREAL, Serie Políticas, n.º 8.

Ravela, P. (2003), *¿Cómo aparecen los resultados de las evaluaciones educativas en la prensa?* Santiago de Chile, Grupo de Trabajo sobre Estándares y Evaluación, PREAL.

Ravela, P. (2006), *Para comprender las evaluaciones educativas. Fichas didácticas*, Santiago de Chile, Grupo de Trabajo sobre Estándares y Evaluación, PREAL.

Ravela, P. (2008), "Evaluación y profesión docente en Uruguay. Problemas y propuestas", *Punto. edu*, n.° 11, p. 17, Montevideo, Centro de Investigación y Promoción Educativa y Social (CIPES).

Ravela, P., Arregui, P., Valverde, G., Wolfe, R., Ferrer, G., Martínez Rizo, F., Mariana Aylwin, M. y Wolf, L. (2008), *Las evaluaciones que América Latina necesita*, documento de trabajo n.° 40, Grupo de Trabajo sobre Estándares y Evaluación, PREAL.

Resnick , L. B. y Resnick, D. P. (1992), "Assessing the thinking curriculum. New tools for education reform", en Gifford, B. R. y O´Connor, M. C. (eds.), *Changing assessments. Alternative views of aptitude, achievement and instruction*, pp. 37-75, Boston, Kluwer Academic.

Roca, E. (2008), *Perspectivas de la evaluación*, V Congreso Internacional sobre Dirección de Centros Educativos, Universidad de Deusto, Bilbao.

Roca, E. y Sánchez, R. (en prensa), "Citizens' competences and education for the 21st century. Working and assessing competences in the spanish education system", en CIDREE, *The implementation of key competences. Challenges and opportunities*.

Rojas, C. y Esquivel, J. M. (1998), *Los sistemas de medición del logro académico en Latinoamérica*, Washington D. C., LCSHD Paper Series.

Rodríguez, L. y Puiggrós, A. (2001), *Proyecto Evaluación en la Escuela*, Argentina, Dossier Evaluación de la Calidad.

Ruiz, V., Nirenberg, O. y Brawerman, J. (2000), *Evaluar para la transformación*, Buenos Aires, Paidós.

Ruiz, V., Nirenberg, O. y Brawerman, J. (2003), *Programación y evaluación de proyectos sociales, aportes para la racionalidad y la transparencia*, Buenos Aires, Paidós.

Sadler, R. (1989), "Formative assessment and the design of instructional assessment", *Instructional Science*, n.º 18, pp.118-144.

Santibáñez, L., Martínez, J. F. *et al.* (2007), *Breaking ground. Analysis of the assessment system and impact of Mexico's teacher incentive program "Carrera Magisterial"*, California, Rand Education.

Saranson, S. B. (2003), *El predecible fracaso de la reforma educativa*, Barcelona, Octaedro.

Sauvageot, C. (1997), *Indicators for educational planning, a practical guide*, París, IIEP-UNESCO.

Schmidt, W. y Prawat, R. (2003), "Curriculum coherence and national control of education, issue or non-issue?", *Journal of Curriculum Studies*, 38 (6), pp. 641-658.

Scriven, V. (1991), *Evaluation Thesaurus*, Londres, Sage Publications, 4.ª ed.

Secretaría de Educación Pública (2002), *Las escuelas públicas mexicanas de educación básica. Algunos aspectos a considerar en la autoevaluación escolar*, México.

Shavelson, R. J., McDonnell, L. y Oakes, J. (1989), *Indicators for monitoring mathematics and science education. A sourcebook*, Santa Monica, Ca, Rand Corporation.

Shepard, L. A. (2006), "Evaluation in the classroom", en Brennan, R. L., *Educational measurement*, ACE/Praeger Westport, 4.ª ed.

Simons, H. (1999), *Evaluación democrática de instituciones escolares*, Madrid, Ediciones Morata.

SITEAL (2007), *Informe de tendencias sociales y educativas 2007*, Buenos Aires, IIPE-UNESCO-OEI, disponible en: http://www.siteal.iipe-oei.org.

Solomon, P. G. (2003), *The assessment bridge, positive ways to link tests to learning, standards and curriculum improvement*, Corwin Press, Thousand Oaks.

Stufflebeam, D. y Shinkfield, A. (1993), *Evaluación sistémica. Guía teórica y práctica*, Barcelona, Ediciones Paidós.

Swanson, D., Norman, G. y Linn, R. (1995), "Perfomanced-based assessment, lessons from the health proffession", *Educational Researcher*, 42 (5), pp. 5-11.

Tedesco, J. C. y Tenti Fanfani, E. (2004), "La reforma educativa en la Argentina. Semejanzas y particularidades", en Carnoy, M. (2004), *Las reformas educativas en la década de 1990. Un estudio comparado de Argentina, Chile y Uruguay*, pp. 21-71, Buenos Aires, Akian Gráfica Editora.

Tenti, E. (2001), *En casa de herrero cuchillo de palo. La producción y el uso del conocimiento en el servicio educativo*, Buenos Aires, IIPE-UNESCO.

Tenti, E. (2003), "Los docentes y la evaluación", en *Evaluar las evaluaciones. Una mirada política acerca de las evaluaciones de la calidad educativa*, Buenos Aires, IIPE-UNESCO.

Tenti, E. (comp.) (2008), *Nuevos temas en la agenda de política educativa*, Buenos Aires, IIPE-UNESCO-Siglo XXI.

Tescarolo, R. (2005), *A escola como sistema complejo. A ação, poder e o sagrado*, Brasil, Escrituras Editora e Distribuidora de Livros, Ltda.

Tiana, A. (1996), "La evaluación de los sistemas educativos", *Revista Iberoamericana de Educación*, n.º 10, pp. 37-61.

Tiana, A. (1996), "Tratamiento y uso de la información en la evaluación", *Programa Evaluación de la Calidad de la Educación. Documentos*, vol. I, Buenos Aires, OEI.

Tiana, A. (1997), "Indicadores educativos. ¿Qué son y qué pretenden?", *Cuadernos de Pedagogía*, n.º 256, pp. 49-53.

Tiana, A., *¿Qué variables explican los mejores resultados en los estudios internacionales?*, disponible en: http://www.Oei.org.ar/noticias/Que_variables_explican.pdf.

Tiramonti, G., Dussell, I., Pinkasz, D., Marcalain, G. y Montes, N., *Un estudio de caso de tres jurisdicciones de la Argentina. Grado de conocimiento de la información disponible y su utilidad en los diferentes niveles de gestión del nivel medio de la enseñanza pública*, Buenos Aires, FLACSO.

UE (2006), *Recomendación del Parlamento Europeo y del Consejo de 18 de diciembre de 2006 sobre las competencias clave para el aprendizaje permanente* (2006/962/CE), Diario Oficial de la Unión Europea, n.º 394, pp. 10-18.

UNESCO (2002), *Estudio cualitativo de escuelas con resultados destacables en siete países latinoamericanos*, Santiago de Chile, Laboratorio Latinoamericano de Evaluación de la calidad de la Educación.

UNESCO (2005), *Informe de seguimiento de la Educación para Todos en el Mundo*, caps. 2 y 4.

UNESCO (2007), *Educación para Todos en 2015, ¿alcanzaremos la meta?* París, UNESCO.

Varas González, C. (1984), *Estructura y funciones. Teoría del gobierno municipal*, pp. 164-188, Santiago de Chile, Universidad de Chile.

Vegas, E. y Umansky, I. (2005), *Mejorar la enseñanza y el aprendizaje por incentivos. ¿Qué lecciones nos entregan las reformas educativas de América Latina?*, Washington, D. C., Banco Internacional de Reconstrucción y Desarrollo.

Vigotsky, L. S. (1978*), Mind in society. The development of higher psychological process*, Cambridge, MA, Harvard University Press.

Wheeler, P. y Haertel, G. (1993), *Resources handbook on performance assessment and measurement. A tool for student, practitioners, and policymakers*, Berkeley, CA, The Owl Press.

Wiggins, G. (1998), *Educative assessment. Designing assessments to inform and improve student performance*, San Francisco, Jossey-Bass Publishers.

Žižek, S. (1998), *Porque no saben lo que hacen. El goce como un factor político*, Buenos Aires, Paidós.

Los autores

Elena Martín

Profesora de Psicología de la Educación de la Universidad Autónoma de Madrid. Participó en el diseño y puesta en marcha de la LOGSE en España. Es miembro del Instituto de Evaluación y Asesoramiento Docente (IDEA) y colaboradora del Instituto de Evaluación Educativa de la Universidad Católica de Montevideo. Desde su responsabilidad como consultora de la OEI, coordina el grupo de expertos en evaluación de esta organización.

Felipe Martínez Rizo

Licenciado en Ciencias Sociales por la Universidad de Lovaina. Desde 1974 es profesor de la Universidad Autónoma de Aguascalientes, México, donde fue director general de Planeación y de Asuntos Académicos, decano del Centro de Ciencias Sociales y rector. Ha escrito más de treinta libros y alrededor de ciento cincuenta artículos. Es investigador nacional y miembro de la Academia Mexicana de Ciencias. Fue el fundador y primer director general del Instituto Nacional para la Evaluación de la Educación (2002-2008).

Leda Badilla

Catedrática de la Universidad de Costa Rica, donde ha trabajado durante más de 27 años. Actualmente dirige el programa de posgrado en Evaluación Educativa en dicha Universidad. Es representante propietaria de la Universidad de Costa Rica en el Consejo Superior de Educación. También representa a Costa Rica ante el Consejo Centroamericano de Acreditación.

Juan Manuel Esquivel

Profesor de Química, B. c. en Educación, B. c. en Química y licenciado en Ciencias de la Educación (Universidad de Costa Rica). M. Sc. en Evaluación e Investigación Educativa y Ph. D. en Currículo e Instrucción de la Universidad de Kansas (Estados Unidos). Profesor de educación secundaria (1965-1978) y de la Universidad de Costa Rica (1970-2006). Catedrático desde 1986. Director de la Escuela de Formación Docente (1976-1980) y del Instituto de Investigación en Educación (1980-1989), y decano en la Universidad de Costa Rica (1989-1992). Consultor de organismos internacionales, director de los proyectos educativos CECC/SICA (2000-2008), autor de artículos y libros.

Guillermo Ferrer

Magíster en Administración y Política Educativa, y en Desarrollo Curricular y Tecnología Educativa. Investigador y consultor en políticas educativas para América Latina y el Caribe, especializado en sistemas nacionales de medición y evaluación educativa, estándares y desarrollo curricular. Docente en programas de posgrado, formador de directivos y asesor en proyectos de redireccionamiento académico.

Marta Kisilevsky

Máster en Ciencias Sociales con orientación en Educación de la FLACSO, y doctoranda en Ciencias Sociales de la misma institución. Autora de libros y artículos publicados en revistas nacionales e internacionales. Ha presentado sus producciones en foros académicos dentro y fuera del país. Fue directora de Programación Educativa de la Secretaría de Educación del gobierno de la

Ciudad de Buenos Aires. Es investigadora y docente en universidades públicas y privadas. Actualmente es directora nacional de Información y Evaluación de la Calidad Educativa del Ministerio de Educación de Argentina.

Álvaro Marchesi

Catedrático de Psicología Evolutiva y de la Educación en la Universidad Complutense de Madrid. Director general de Renovación Pedagógica y secretario de Estado de Educación desde 1986 a 1996. Coordinador del Instituto de Evaluación Educativa (IDEA) desde 1997 a 2006. Sus tres últimos ensayos son: *Controversias en la educación española, ¿qué será de nosotros, los malos alumnos* y *Sobre el bienestar de los docentes. Competencias, emociones y valores.* En la actualidad es el secretario general de la Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI).

Patricia McLauchlan de Arregui

Magíster en Sociología, egresada de la Universidad Católica del Perú y de la Universidad Johns Hopkins de Baltimore. Trabaja en GRADE, centro peruano de investigación sobre políticas económicas y sociales, del cual fue directora ejecutiva entre 1988 y 1998. Ha estudiado aspectos diversos de la educación superior y la formación docente, así como sobre la educación básica y la evaluación de logros de aprendizaje. Coordina el Grupo de Trabajo sobre Estándares y Evaluación del PREAL.

Margarita Poggi

Directora de la Sede Regional Buenos Aires del Instituto Internacional de Planeamiento de la Educación de la UNESCO (IIPE/UNESCO). Ha sido profesora en diversas maestrías en distintas universidades y ha coordinado equipos de investigación educativa. Se ha desempeñado como directora nacional de Evaluación en el Ministerio de Educación de Argentina y como directora general de Planeamiento en el gobierno de la Ciudad de Buenos Aires Es autora de numerosas publicaciones sobre innovación educativa, evaluación, formación de directivos e instituciones escolares.

Pedro Ravela

Profesor de Filosofía en educación media (Instituto de Profesores Artigas, Uruguay) y magíster en Ciencias Sociales y Educación (FLACSO, Buenos Aires). Investigador en la CEPAL (1990-1995), director de la Unidad de Medición de Resultados Educativos (ANEP, 1995-2001), coordinador nacional del estudio PISA en Uruguay (ANEP, 2002-2006) y coordinador técnico del LLECE (UNESCO, 2005-2006). Profesor y consultor en Evaluación Educativa en distintos países de América Latina. Actualmente dirige el Instituto de Evaluación Educativa en la Universidad Católica del Uruguay.

Enrique Roca

Director del Instituto de Evaluación desde septiembre de 2006. Ingeniero Técnico Topógrafo, licenciado en Geografía e Historia, profesor de Historia Económica Mundial y catedrático de Bachillerato. Ha sido director del Centro de Investigación y Documentación Educativa del Ministerio de Educación y Ciencia (España), asesor del secretario de Estado de Educación y miembro del Instituto IDEA. Representante español en el Comité de Educación de la OCDE y en la IEA. Ha publicado numerosos trabajos de historia económica, de educación y de evaluación.

Alejandro Tiana

Doctor en Filosofía y Letras (Pedagogía) por la Universidad Complutense de Madrid. Catedrático de Teoría e Historia de la Educación en la Universidad Nacional de Educación a Distancia (UNED). Entre 2004 y 2008 fue secretario general de Educación en el Ministerio de Educación y Ciencia. En la actualidad dirige el Centro de Altos Estudios Universitarios de la OEI. Es autor o coautor de 17 libros y más de cien artículos o capítulos de libros sobre diversos temas educativos.